Development of a traffic model for predicting airborne PAH exposures since 1960 on Long Island, New York.

Report to the National Cancer Institute, and the
National Institute of Environmental Health Sciences
for work completed under USPHS Grant U01-CA/ES-66572

FINAL DRAFT
July 1, 2005

Jan Beyea[1], Maureen Hatch[2,3] Steven D. Stellman[4], Susan L. Teitelbaum[2], Marilie D. Gammon[5]

1: Consulting in the Public Interest (CIPI) Lambertville, NJ, 08530, jbeyea@cipi.com;
2: Department of Community and Preventive Medicine, Mount Sinai School of Medicine;
3: currently, Division of Cancer Epidemiology and Genetics, National Cancer Institute, DHHS;
4: Department of Epidemiology, Mailman School of Public Health, Columbia University;
5: Department of Epidemiology, University of North Carolina at Chapel Hill

Abstract:

We have developed an historical reconstruction method for estimating past exposure to airborne PAH for use in epidemiological studies in which it is required to estimate PAH exposure over long periods of time. Our model has been developed specifically for use in a population-based, case-control study of breast cancer in relation to environmental factors on Long Island, NY, but the methodology is generalizable to any geographic area for which suitable environmental data are available. Methods: We used benzo[a]pyrene as a surrogate for total PAH exposure because the early data in the literature were largely limited to BaP. We modeled traffic PAH exposures because prior studies indicated that traffic PAH dominates ambient concentrations in the northeastern US. To estimate traffic emissions over time and space we relied on quantitative histories of both tailpipe emissions in μg per vehicle-km previously developed for this study and the number of vehicles moving on individual km-road segments (taken from historical county traffic maps or tabulations). An emission model for engines during the cold, warm-up phase was also developed. Standard meteorological dispersion and deposition models were used to translate emissions in μg per km per day along the NY metropolitan area road network into predicted air concentrations at a study subject's residence over the years. For use in model validation and calibration, the model predicts residential soil and carpet dust concentrations, as well as the contribution of traffic PAH to levels of PAH-DNA adducts in blood. Results: Individual exposures showed strong peaks in the 1970's, strong variation with spatial location (especially near traffic intersections), and large variation in cumulative exposure across study subjects matched by age of arrival on Long Island. Exposure opportunity indexes (EOIs) based on the geographic model are suitable for estimating an individual woman's exposure during specific calendar years as well as susceptible age periods. Our geographic modeling approach complements exposure markers based on environmental samples and biomarkers of PAH damage which tend to reflect recent exposures. It also complements markers of PAH exposure derived from questionnaire data on diet and smoking history.

Introduction

There is evidence that PAH may play a role in the causation of breast cancer (El-Bayoumy 1992), (Li, Zhang et al. 1999), (Morris and Seifter 1992), but estimation of individual PAH exposure is a challenge, especially historical exposure. Exposure to PAH occurs via several routes, including air, water, and food. We present here details of a method of estimation of historical exposure from airborne sources that is intended for use in epidemiological studies in which geocoded data (latitude and longitude) can be obtained for residential and occupational locations of study participants.

Our work was undertaken as part of the, “Long Island Breast Cancer Study Project” (LIBCSP) (Gammon, Neugut et al. 2002), a case/control study involving approximately 3000 women on Long Island, NY. The study area is shown in Figure 1. The present paper describes the historical reconstruction method and resulting geographic model, which is a specific example of a generic approach described in an earlier paper (Beyea and Hatch 1999).

Although historical exposure models can provide individualized estimates of high specificity, they can involve many parameters known imprecisely or only by inference. Consequently, in our development of exposure estimates, we have also stressed model validation and calibration, which will be the topic of a subsequent paper. Of particular interest in our methodology is the ability to account for the higher emissions that occur at intersections, known to be emission hot spots (Sheu, Lee et al. 1996a; Sheu, Lee et al. 1996b), (Sculley 1989).

Air exposure appeared to be the best candidate for historical reconstruction of PAH. As summarized in (Eder 1999), studies to date have indicated that environmental exposure to airborne PAH can have a larger effect on blood PAH-DNA adduct levels than industrial exposure, food or smoking (Eder 1999), in spite of the fact that, on a strict mass-balance basis, the amount of PAH entering the body via the air is estimated to be approximately 5-10% on average of the dietary contribution (Lioy, Waldman et al. 1988), (Venkataraman and Raymond 1998), (Naumova, Eisenreich et al. 2002), (Gammon, Santella et al. 2002).

Within the category of PAH air exposure, we focused on traffic emissions because they are a major source of both indoor and outdoor exposures to PAH, and often the largest source in areas near cities, as has been confirmed in a number of experimental studies (Dubowsky, Wallace et al. 1999), (Lim, Harrison et al. 1999), (Harkov, Greenberg et al. 1984), (Dickhut, Canuel et al. 2000), (Dunbar, Lin et al. 2001), (Levy, Houseman et al. 2001). An earlier case-control study of breast cancer on Long Island reported an elevated (but not statistically significant) odds ratio for breast cancer in areas of high traffic density (Lewis-Michl, Melius et al. 1996). Elevated risk was also reported in areas with two or more chemical facilities.

Indoor sources of PAH can also contribute to an individual's total air exposure, especially in episodic events, including smoking (Sakai, Siegmann et al. 2002), (Mitra and Ray 1995), and for 3-ringed PAH compounds (Naumova, Eisenreich et al. 2002). However, studies by Sheldon et al. have found outdoor BaP highly correlated with indoor levels in California, contributing more than 50% on average to indoor levels (Sheldon, Clayton et al. 1992). Other studies have found similar results: The contribution of outdoor sources to the heavier PAHs was 63%-80% for 5-7 ringed compounds in US cities (Naumova, Eisenreich et al. 2002)) and 76% for BaP in Japanese cities (Ohura, Amagai et al. 2004)).

Our estimates of historical exposure complement short-term PAH biomarker measurements carried out in the same population using PAH-DNA adducts, which are indicative of recent DNA damage (Gammon, Santella et al. 2002). In this population, an ~50% increase in breast cancer risk was noted in relation to PAH-DNA adducts found in peripheral blood. No trend in risk with number of adducts was observed, which could be interpreted most simply as an indication that the dose-response for PAH is non-linear. Alternatively, these findings may suggest that individual differences in the response to similar levels of PAH exposure may be more relevant in breast carcinogenesis (Gammon, Santella et al. 2002), (Dickey, Santella et al. 1997).

PAH-DNA adducts reflect only recent exposures; therefore, it is of interest to explore PAH exposure in the more distant past, which could be more important for breast carcinogenesis.

Because biomarkers are not available that reflect exposures in the distant past, we have turned to geographic modeling.

Methods

We present here a methodology that we have developed for estimating historical exposure to traffic-source PAH within the context of a geographic information system. The methodology utilizes a modeling procedure that builds on a variety of traffic-related parameters for which substantial data resources exist over a long time period. The ultimate goal is to model the exposure for each study participant, based upon a lifetime history of her residential and possibly occupational locations.

The methods used in this study to develop estimates of cumulative PAH exposure are most similar to those used in an estimate of historical NOX- and SO2-exposure as part of a case/control study of lung cancer in Sweden (Bellander, Berglind et al. 2001) and to recent modeling work carried out in Portland, Oregon (Cohen, Cook et al. 2005) and earlier in Denmark (Raaschou-Nielsen, Hertel et al. 2000). The methods are analogous to those developed many years ago to predict carbon monoxide concentrations from traffic networks (Johnson, Ludwig et al. 1973). We extend the methodologies in these earlier studies to account for the distinct, geospatial exposure patterns produced by emissions at intersections and during engine warm-up (cold-engine emissions). Emissions were determined first for warm engines (cruise conditions). Next, emissions were adjusted relative to cruise emissions for the cold-engine segment of trips and for acceleration and deceleration at traffic intersections. To estimate exposures to residential and work locations, we first estimated the distribution of emissions per km along the NY metropolitan area road network, which runs through 22 counties in New Jersey, New York State, and Connecticut, and then used standard meteorological dispersion and deposition models to translate these emissions, in units of μg per day per km, into predicted air concentrations (ng/m^3) at a study subject's residence or work location over the years. For purposes of model validation and calibration, we made predictions of quantities closely related to PAH air concentration for which field data could be obtained. (PAH air-concentration measurements were not available for direct model validation and calibration.) These secondary predictions included carbon monoxide concentrations at USEPA monitoring stations, which are dominated by traffic emissions in urbanized areas (Chang and Weinstock 1973), (USEPA 2003). We also made predictions for concentrations of PAH in soil and carpet dust at a study subject's residence, as well as the contribution of traffic PAH to levels of PAH-DNA adducts in blood. Each of these predictions is compared with samples collected for the LIBCSP (Gammon, Neugut et al. 2002). We estimated traffic emissions over time and space using quantitative histories of both tailpipe PAH emissions per vehicle-km and the number of vehicles moving on individual km-road segments.

We took B[a]P as a surrogate for PAH, because it is considered a good marker of overall PAH exposure (Fertmann, Tesseraux et al. 2002) and the best single marker of PAH carcinogenicity (Hannigan, Cass et al. 1998). Furthermore, B[a]P was the only PAH reported consistently in the historical literature..

To obtain indoor exposures, estimated outdoor PAH concentrations were multiplied by an average building penetration factor of 0.75 (Long, Suh et al. 2001). The impact of random variations about this value for each study subject was studied, as discussed below in the section on sensitivity analysis.

Exposures were computed for 1960, 1970, 1980, 1990 based on the traffic counts or number of cold starts for those years. Exposures for years between 1960 and 1990 were obtained by interpolation between the appropriate exposure pair. Exposures after 1990 were obtained by logarithmic extrapolation from 1990. Exposures before 1960 were obtained by logarithmic extrapolation backwards from 1960.

Traffic flow by location, year, and engine operating condition

Table 1 lists the data sources and methods that were used to derive the traffic flow on the network, broken down by cruise, cold-engine and intersection components. Data sources for flows of warm engines included over 13,000 measurements of “annual average daily traffic (AADT)” recorded on paper maps or lists back to the 1960s. The median distance to the nearest traffic count on major roads in Long Island was 0.6 km, with a mean distance of 1.4 km. (Figure 2 gives an idea of the density of available measurements.)

Cold-engine emissions differ from warm-engine emissions both spatially and by time of day. Data sources used to develop traffic flows for cold engines included number of households at the census block level and extracts from the travel-diary database of the National Personal Transportation Survey (NPTS). The travel-diary extracts gave us the number of cold starts per household per hour of the day for our study area (USDOT 1996). To determine the fraction of the traffic flow that accelerated or decelerated at an intersection, the AADT were corrected to account for the fraction of vehicles that had to slow down or stop due to traffic controls. A simplified version of models used to predict peak CO exposures at intersection was used (Nelli, Messina et al. 1983).

Relative emissions by hour of the day for warm-engine emissions were taken as proportional to hourly traffic counts, as represented by a typical traffic density curve measured in Nassau County (NCOPT 1978). See Figure 3. This curve is consistent with other measurements (Cardelino 1998). The time-of-day curve for cold-engine emissions was set equal to the hourly frequency of trip starts on Long Island, as extracted from (USDOT 1996). It is similar to the curve for cruise emissions, except for a reduction in cold-starts in the afternoon. It is also shown in the figure.

Emissions

Emissions per km of roadway per day are a product of traffic count (vehicles per day) and tailpipe emissions per vehicle (μg/km). Determination of the historical variation in tailpipe emissions per km has been determined in previous work (Beyea, Hatch et al. 2005) based on measurements carried out in tunnels or on individual vehicles run in place on test beds (dynamometers). The results are reproduced in Figure 4. Emissions were further adjusted to account for non-cruise operating conditions, such as cold-engine operation or acceleration and deceleration at traffic intersections.

Digital street maps are made up of small straight-line segments strung together—some 500,000 in the road network considered in our study. For a given engine operating condition, emissions per from one of these straight-line road segments was set equal to product of the AADT at the segment, the length of the segment, and the appropriate BaP emitted per vehicle-km in μg/km.

Total emissions were written as the sum of cruise, cold-engine, and intersection emissions. For the default model, the coefficients of the various components were estimated from the literature as reported in (Beyea, Hatch et al. 2005) and summarized in Table 2.

As indicated in Table 2, it is appropriate to scale emissions near intersections by a factor of ten, but given the paucity of data, the sensitivity of any results to this parameter should be explored or measured during model calibration. In our model, increased emissions are graded with distance from the intersection in three steps, with the full scale factor applied within 12.5 meters and reduced by 1/3rd from 12.5 to 25 meters. The scale factor was reduced by 2/3rds between 25 and 50 meters. The sensitivity of final results to these relative reduction factors can be explored through sensitivity analyses or model calibration. One such validation and optimization effort for this and the model parameters discussed below will be reported in a subsequent paper.

Data on the increase in emissions during cold start is more plentiful than data on increases at intersections. The average scale factor is 8.0. However, the uncertainty range is large, which suggests that this scale factor also should be validated and optimized using field measurements. The default travel time necessary for an engine to warm up was taken as 1 km (Ahlvik, Almen et al. 1997).

Data on seasonal differences in emissions is limited. As discussed in (Beyea, Hatch et al. 2005), we doubt winter tailpipe emissions differ by more than a factor of two in the Long Island area. This is a small difference compared to the historical change over time in emissions shown in Figure 4. Furthermore, the increase would occur during cold starts, so falls under the domain of the cold-start model, which itself has been assigned an average scale factor of 8 in the default model. On an annual basis, accounting for increased emissions during winter cold starts, will only change the average cold-start scale factor. Since the cold-start scale factor has to be varied anyway for sensitivity analyses or estimated through model calibration, there is no need to take separate account of winter/summer emissions for annual exposure estimates. Study subjects who moved during a particular year could have part of their exposure from cold starts miscalculated for that year, but we expect this effect to be very small.

Background emissions

Previous traffic models have used gasoline sales data to estimate background emissions or have used a constant value as a surrogate for the cumulative contribution of distant roads (Viras, Siskos et al. 1987), (Raaschou-Nielsen, Hertel et al. 2000). Although we have attempted to directly account for traffic emissions within 80 km of our study area, there are more distant roads, as well as other sources of PAH emissions, making it prudent to include a background term in the model. We have provided two options for use in model calibration exercises. In the first, a constant term is added to the model prediction, while the second option makes background proportional to the exposure calculated from the more distant counties (all but Nassau, Suffolk, and Queens counties). In both approaches, the relative scale of the background term can be set by calibration to environmental measurements.

Meteorological dispersion model

The contribution from each point source to air concentration at a downwind receptor (Chock 1978) was computed within 100 meters of a road using a highway line-source model applied to each of the 500,000, straight-line road segments in the traffic network. The “Chock” highway model was chosen because it gave the best fit, when compared to a suite of models tested, to tracer concentrations near the Long Island Expressway as part of a test carried out by the New York Department of Environmental Protection (Sistla, Samson et al. 1979). R^2 values ranged from 0.75 to 0.92 for various meteorological conditions and angles to the road.

Beyond 100 meters, we used a standard, Gaussian puff dispersion model (equivalent to the USEPA's "RAM" model (Catalano, Turner et al. 1987)). Table 3 describes the default parameters and data sources used in the dispersion model.

Total concentration at a study subject's residence in units of ng/m^3 was computed as the sum of the contributions from the approximately 500,000 source segments.

Meteorological models require data on wind speed, direction, and other variables that govern the movement of puffs of emitted pollution. Hourly meteorological data collected at Brookhaven National Laboratory in Suffolk County Long Island in 1993 were used for all traffic segments and all years. Complete data were not available for every year of interest in our study and it would have been computationally prohibitive to account for year-to-year variations in meteorological parameters. Meteorological data from another year (1990) and location (MacArthur airport) were used in sensitivity tests (results discussed below).

Model Output

The output of the model has been designed to facilitate comparison with environmental data to be used in validation and calibration exercises. In addition to providing air concentrations, the model also provides estimates of PAH soil concentrations, based on dry and wet deposition, as well as estimates of PAH in carpet dust per m^2, based on dry deposition. Finally, the model can be run with all plume depletion processes turned off, which produces output proportional in any hour to output from a traffic-generated, carbon monoxide model.

Application of Model to Study Population:

Geocoding

The model was used to develop exposure estimates for 3064 participants in the Long Island Breast Cancer Study Project. A total of 1508 cases and 1556 controls completed the interviewer-administered main questionnaire (http://epi.grants.cancer.gov/LIBCSP/projects/Questionnaire.html). Residential addresses were obtained from each subject for all homes in which she resided one year or longer in Nassau and Suffolk counties. Occupational addresses were also sought for all jobs held six months or longer (Gammon, Neugut et al. 2002).

Digital street maps were purchased from BLR DATA, Inc. (now part of GDT, Lebanon, NH). Addresses were geocoded using BLR software, with manual cleaning of poorly coded addresses. Two levels of geocoding success were carried forward into the calibration process. For the less accurate level, we allowed extrapolation of address number from the start of a dead-end street. For the more accurate level, we excluded residences on dead-end streets, which allowed interpolation of address numbers between known values on either side of the geocoded location.

Exposure opportunity Index (EOI)

Residential exposure calculations assume a women is in her residence whenever she is not at work, and work exposures assume she is at her workplace for the number of hours per week recorded in the interview. Thus, our calculated exposures are exposure opportunities which could be incorporated into toxicological measures that also model breathing rate, metabolic activity, and other physiological parameters

The model can accommodate both residential and workplace exposures by introducing assumptions about the relative time spent by a subject in each location, but its accuracy is severely limited by the quality of workplace address reporting. For example, only 10% of work addresses given at interview would fully geocode to the street level, although almost all were determinable at the city or town level. Thus, it was not possible in this study to include an exposure index computed down to the level of individual work locations. To see if group level exposure markers at the city or town level might prove useful, we prepared a work exposure index imputed for each town or city for each month from 1960 to 1997. Such an index would serve at least to differentiate exposures in the more urbanized towns from the more rural villages on Long Island. GIS methods were used to assist in the imputation of work exposures.

Hypothetical workplaces were assigned along the major roads in the town or city in which a woman worked. EOI's were computed for each hypothetical location and averaged to obtain an imputed EOI.

Results

Geocoding success

The total number of residential locations recorded at interview was 8321. The maximum number of Long Island residences reported by a woman was 17 (median = 2). Only geocodes to the street level were counted as successful, which meant excluding the latitude and longitudes for zipcode centroids that GIS programs produce, when more accurate determinations cannot be made. Statistics for the address locations are given in Table 4. 96.4% of locations had a street name recorded, but only 79% of the streets had accompanying street numbers, which is a requirement for street-level geocoding. Of the locations with street numbers, 87% geocoded to the street level. The overall geoding success rate for residential addresses by year of residence is shown in Figure 5. The overall success rate was 85% for current addresses, 65% for 1960 residences, and in the 30 to 40 percent range before 1940. We note that accuracy in determining latitude and longitude of current residences, but not former residences, could be improved in future studies by taking to interview hand-held GPS units, which are now quite accurate and inexpensive.

A total of 14,668 work addresses were recorded at interview (median 4 per subject). Approximately half were missing a street name. Of the remaining addresses, only 4532 were in the study area and had street names, and many lacked street numbers.

Spatial and temporal variation of exposures

In 1996, 50% of the study population lived within 250 meters of one of the major roads on the traffic network and 80% lived within 500 meters. The spatial pattern of emissions from intersections is dominated by cross-like structures as shown in Figure 6. The intersection segments shown in the figure extend 100 m from the actual intersection.

The predicted air concentration along a perpendicular transect across Long Island is shown in Figure 7. Clearly, PAH air exposures are highly variable in space, depending on proximity to a major road. The location of the transect is indicated on the map in Figure 1.

Year-by-year exposure estimates are shown in Figure 8 for a hypothetical study subject to demonstrate several important features of typical exposure profiles. Yearly exposures change dramatically because of shifts in residential locations and because of temporal variations in tailpipe emissions. There are also gaps in the exposure sequence due to 1) absence from Long Island, 2) residence on Long Island for less that the one year required by the questionnaire protocol, or 3) an address that did not accurately geocode to the street level.

Sensitivity tests

Exposure estimates depend on a number of uncertain parameters and assumptions, whose impacts on quantities of epidemiological interest can be examined through sensitivity and uncertainty analysis. Uncertainty analysis, which can easily incorporate correlations between uncertain parameters, is best addressed after model calibration through Monte Carlo techniques. (USEPA 1996). Sensitivity analysis, which generally looks at changes in model output as one parameter is varied at a time (Morgan and Henrion 1990), can be done before calibration. For this report, we perform sensitivity tests on an indicator that reflects averages over the study population. We chose to look at variations in population averages for practical reasons, because the model predicts thousands of individual exposure estimates that are correlated in complex ways. Furthermore, estimates of means, especially differences in means between samples representing cases and controls, is an appropriate test of a model that is to be used in an epidemiological analysis of exposure and disease. If the model uncertainty associated with the differences in means between hypothetical cases and controls is much greater than the corresponding sample-size uncertainty, then the model is unlikely to be very useful. To investigate the changes in averages over study populations as model parameters and assumptions were changed, we produced 40 “exemplary data sets” (O'Brien and Muller 1993), (O'Brien 1998), each of which consisted of a randomly selected, split sample of the study population. For each of the replications, we computed population averages for the two split populations and then looked at the difference. The root-mean-square differences of the 40 split samples was then computed and standardized to percentage format. No distinction was made between cases and controls.

Table 5 presents the results of the analysis of the percentage difference in cumulative exposure of split samples of women in the study. The percentage differences arising from individual model uncertainties are all smaller than the percentage difference that occurs from sample size variations alone, which is 5.5%.

If we combine the uncertainties presented in Table 2 by taking the square root of the sum of squares, the total variation in the group mean exposure amounts to 3.5%, which is less than the sample size variability of 5.5%. Such an estimate of combined uncertainty is not completely rigorous, however, because there is some degree of correlation between some of the variables (e.g., deposition velocity and penetration rate of fine particles). However, the total model uncertainty would only rise to 8%, assuming all the variables in Table 2 were 100% correlated in one direction. In fact, it can be expected that some variables are uncorrelated and some will have compensatory correlations. For these reasons, it seems unlikely that model uncertainty is out of line with the sample size variability. Model uncertainty should not dominate the epidemiologic analysis for sample sizes in our study population, although this conclusion needs to be checked with a post-calibration Monte Carlo analysis.

Cumulative exposure opportunity

The exposure model produces monthly average exposure estimates, which can be summed for various time periods of interest. First, we compute a Long Island EOI for a woman from 1960-1990, the longest period for which we have reliable emissions and traffic pattern information. Next, we compute exposures for two time periods when the breast is thought to be particularly susceptible to carcinogenic insult (Colditz 1995), (Colditz and Frazier 1995): cumulative exposure to age 20 and, cumulative exposure to age at first birth. If a woman was born or arrived on Long Island before 1960 then the cumulative exposures begin with 1960. Otherwise, exposures begin with time of birth or arrival on Long Island.

The range in cumulative EOI (CEOI) is substantial. For women present on Long Island for the 30 years from 1960 to 1990, the ratio of maximum to minimum CEOI is 175, assuming default model parameters. The ratio of the 90th percentile CEOI to the 10th percentile CEOI is a factor of 4. The logarithms of the cumulative exposure distributions have an approximate normal distribution, with a geometric standard deviation of 1.9.

Missing exposure data before arrival on Long Island.

Missing data can occur for a number of reasons. Consider first the period before a study subject moved to Long Island. The study design did not include gathering information on residences before the year of arrival on Long Island; nor would it have been feasible to estimate traffic exposures at the street level throughout the country. Therefore, we have no way of making exposure estimates off Long Island, which means that we are missing all exposures from 1960 to arrival year for those women who moved to Long Island after 1960. Since tailpipe emissions were significantly higher in the early years all over the country, we could potentially be missing significant exposures, if a woman arrived on Long Island after 1980.

Thus, our cumulative exposure computations are limited to "years of calculable exposure" (YCE). In comparing calculated exposures to breast cancer risk, it will be important to account for YCE. Data are not available to estimate exposures for women during their off-island period, so we have considered two alternatives First, at the suggestion of Sylvan Wallenstein (personal communication, 2003), we have computed relative exposures within groups of women arriving around the same time on Long Island and then pooled the relative exposures across year-arrived groups.

As an alternative, we have also computed cumulative exposure for a nested set of time periods, which begin at different decades, but end with the 1990 exposure stop date. For these computations, we require residence on Long Island to have begun prior to the start date. In this way, we force the years of calculable exposure to be the same for every woman in the individual comparisons, although the set of women may change from time period to time period.

Missing exposure data after arrival on Long Island

Street-level latitude and longitude may be unavailable because an address is incomplete or will not otherwise fully geocode. Also, there may be no address given for a period of time, either because the woman did not live on Long Island during the interval or the residence duration did not meet the one-year minimum specified by the study design. We also did not compute an exposure for years when duration for two residences overlapped in time and the conflict could not be resolved as a dual use of a summer residence. Dual residence was only allowed if one of the locations could be assigned to a known summer district, e.g., Fire Island, Southold, East Hampton. In those 30 cases, exposures were split between the residences, with the summer residence weighted one-third and the other residence two-thirds.

Table 6 shows how the number of women available for analysis changes for different constraints on the percentage of missing data allowed after arrival on Long Island. Table 7 shows how the statistics break down, if we require women to have moved to Long Island before the start year. Note that a woman may have an acceptable estimate for one time period, e.g., 1960-1990, but not have one for another time period, e.g., birth to age 20. A scatter plot of the 1960-1990 cumulative exposures by year of a woman's arrival on Long Island is shown in Figure 9.

Discussion

During the historical period for which we have reliable emissions data, as well as for periods when a woman may be most susceptible to carcinogenic insult, we have produced individualized exposure opportunity estimates for airborne traffic PAH. Our model accounts for historical changes in vehicle emission and traffic flows, and for emissions near intersections and during engine warm-up. The individual exposure estimates generated by the model show distinct patterns by time and space. Estimated PAH concentrations peak in the 1970s, prior to the introduction of catalytic converters. There is considerable spatial variation in PAH concentrations, with peaks near traffic intersections. In addition, estimated cumulative exposure levels vary widely, even among subjects matched on age of arrival on Long Island.

The exposure model has been built to facilitate validation exercises and parameter calibrations. The method can be adapted to other similar locations. Some limitations should be noted, however.

Row housing. Since the percentage of row housing is small in the study area, we did not incorporate "canyon effects" (Raaschou-Nielsen, Hertel et al. 2000) into the modeling.

Sea breezes. Although we included the effects of sea breezes indirectly through our use of Long-Island specific meteorological data, which change in response to sea breezes, we did not explicitly model them. This means we explicitly omitted consideration of the "thermal internal boundary layer" (Luhar and Sawford 1996) that forms over the Island for several hours when the wind is blowing from the ocean on warm days.

Historical changes in road network. From an analysis of area maps, we found that, with few exceptions, all major roads in Nassau and Suffolk Counties were in place by 1960, so we did not include an algorithm to “remove” road sections from the network in backwards extrapolation. This means that in a few cases (e.g., extension of the Long Island Expressway), we will overestimate exposures in the early years due to the inclusion of emissions from roads not yet built (phantom source terms). We do not expect the *inclusion* of a few phantom roads to show any more of an impact on the epidemiologic analysis than the negligible impact we found for the *omission* of 1% of the road segments because of lack of traffic data.

In-vehicle exposures. We cannot account for individualized exposure to traffic PAH while driving (Ott, Switzer et al. 1994), (Brice and Roesler 1966), because no questions directly relevant to in-vehicle modeling were included in the study questionnaire. From examination of the concentrations predicted along the transect crossing Long Island (Figure 7), as well as a comparable curve for Suffolk county not shown, we conclude that peak exposures on roads were generally ten times higher than average residential exposures. Thus, assuming 1 hour of vehicle traffic per 24-hour day, exposures while driving would contribute 10/24ths of average model exposures, leading us to conclude that travel-time exposures did not dominate total exposure for the overwhelming number of study subjects. On the other hand, the omission of in-vehicle exposure may be significant for women living in locations with very low exposure from general traffic. If so, the omission would add some non-linearity between our exposure estimates and the values that would have been computed with a model capable of tracking individualized, in-vehicle exposures.

Inability to model work exposures. As noted earlier, very few women provided work address information sufficient to allow street-level geocoding. Imputation to the town or city level was necessary to generate cumulative exposure estimates. In the future, questionnaires could be designed so that the company name of the employer is requested, which would allow independent assignment of the work location.

Omission of other PAH air exposures. Our study is premised on the experimental finding that traffic exposures are a dominant source of airborne PAH, both in and outdoors. We did not consider a number of additional sources of potential PAH exposure, such as emissions from space-heating furnaces and indoor sources (e.g., PAH in airborne combustion products of cigarette smoke and cooking) Data on home heating sources would have to be obtained through interviews or from records.

<u>Limitations due to incomplete exposure data</u>. When computing cumulative EOIs for a woman, multiple addresses are often needed, some going back quite far, which means that there is a significant chance that a woman may have at least one exposure gap, particularly if the starting year of the cumulative exposure is set earlier than 1960. Success rate for a cumulative EOI depends on the start year and the tolerance level that is set for missing exposures. This makes it especially problematic to take exposure back to childhood for older women. The choice of start year is a tradeoff between increasing the exposure duration and decreasing the precision of the exposure estimates. Our confidence in both the historical traffic flow rate and the tailpipe emission rate drops off dramatically prior to 1960. The geocoding success rate also falls off sharply for residences occupied before 1960. For these reasons, we chose1960 as our starting exposure year.

Conclusion

Although relatively new in environmental epidemiology, geographic exposure modeling combined with historical reconstruction has been used in a number of epidemiological studies of large populations and to assess exposure to environmental hazards (Bellander, Berglind et al. 2001) , (Ward, Nuckols et al. 2000), (Gunier, Harnly et al. 2001), (Brody, Vorhees et al. 2002), (Stellman, Stellman et al. 2003), (Reynolds, Von Behren et al. 2003).   When exposures may have occurred in wide geographical areas over a period of many years, geographic modeling methods may be the only practical exposure methodology available. Nevertheless, confidence in modeling approaches will be strengthened by careful attention to methods. Whenever possible, predictions should be made of quantities related to a model’s exposure index that can be validated against measurable data. Reconstruction of exposures from traffic provides opportunities for improving GIS methods, because there is a relative abundance of information available for both constructing EOIs and validating them.

Acknowledgements

Supported by USPHS Grant U01-CA/ES-66572 from the National Cancer Institute and National Institute of Environmental Health Sciences (J.B., M.H.). Other USPHS grants providing support: CA-63021, CA-72562, CA-17613, and CA-68384 (S.D.S). Also, P30ES10126 from the National Institute of Environmental Health Sciences (M.D.G.). In addition, we received a grant from a group of foundations: Long Island Breast Cancer Coalition, Babylon Breast Cancer Coalition, West Islip Breast Cancer Coalition for L.I., Inc., Huntington Breast Cancer Action Coalition, Noreen T. Holland Breast Cancer Foundation, Inc., Breast Cancer Grassroots Organizations for a Unified Purpose, Inc. The following individuals provided information and/or assistance: James Cook. Laura Palmer, Suffolk County Department of Public works, Michael J. Finocchi, Nassau Department of Public Works, Ed Finster, USEPA, Roy Fedelem, Long Island Regional Planning Board, Anthony Dentrone, Hagstrom maps. We also thank Dr. Iris Obrams (at NCI) for support of our pilot work and Dr. Gwen Collman (at NIEHS) for overall support.

References

Ahlvik, P., J. Almen, R. Westerholm and D. Ludykar (1997). Impact of a block heater on regulated and some unregulated emissions from a gasoline fueled car at low ambient temperatures. SAE International Fall Fuels & Lubricants Meeting & Exposition, Tulsa, Oklahoma, Society of Automotive Engineers, Warrendale, PA, SAE Technical Paper 972908, 1997.

Bellander, T., N. Berglind, P. Gustavsson, T. Jonson, F. Nyberg, G. Pershagen, et al. "Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm." Environ Health Perspect 2001; 109:633-9.

Beyea, J. and M. Hatch. "Geographic exposure modeling: a valuable extension of GIS for use in environmental epidemiology." Env. Health Perspectives 1999; 107, Supplement I:181-190.

Beyea, J., M. Hatch, S. D. Stellman and M. D. Gammon. "Emissions of polycyclic aromatic hydrocarbons from US gasoline-powered vehicles from 1960 to 1995 for use in epidemiologic studies." Submitted. 2005.

BLR. Traffic Count Files for the NYMSA. Lebanon, NH, Business Location Research (purchased by GDT) 1997. 1997.

Brice, R. and J. Roesler. "The exposure to carbon monoxide of occupants of vehicles moving in heavy traffic." JAPCA 1966; 16:597-600.

Brody, J. G., D. J. Vorhees, S. J. Melly, S. R. Swedis, P. J. Drivas and R. A. Rudel. "Using GIS and historical records to reconstruct residential exposure to large-scale pesticide application." J Expo Anal Environ Epidemiol 2002; 12:64-80.

Cardelino, C. "Daily variability of motor vehicle emissions derived from traffic counter cata." J Air Waste Manage Assoc 1998; 48:631-645.

Catalano, J. A., B. D. Turner and J. H. Novak. User's Guide for RAM -- Second Edition. Research Triangle Park, Environmental Protection Agency 1987.

Census_Bureau. 1990 Census of Population and Housing; Population and Housing Unit Counts (US). Washington, Bureau of the Census 1993.

Chang, T. Y. and B. Weinstock. "Urban CO concentrations and vehicle emissions." JAPCA 1973; 23:691-696.

Chock, D. P. "A simple line-source model for dispersion near roadways." Atmos Environ 1978; 12:823-829.

Cohen, J., R. Cook, C. R. Bailey and E. Carr. "Relationship between motor vehicle emissions of hazardous pollutants, roadway proximity, and ambient concentrations in Portland, Oregon." Environmental Modelling & Software 2005; 20:7-12.

Colditz, G. A. "Fat, estrogens, and the time frame for prevention of breast cancer." Epidemiology 1995; 6:209-11.

Colditz, G. A. and A. L. Frazier. "Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus." Cancer Epidemiol Biomarkers Prev 1995; 4:567-71.

CONNDOT. Traffic Log of State Numbered Routes and Roads, 1970, 1980, 1990. Hartford, Connecticut Department of Transportation 1991.

Dickey, C., R. M. Santella, D. Hattis, D. Tang, Y. Hsu, T. Cooper, et al. "Variability in PAH-DNA adduct measurements in peripheral mononuclear cells: implications for quantitative cancer risk assessment." Risk Anal 1997; 17:649-56.

Dickhut, R. M., E. A. Canuel, K. E. Gustafson, K. Liu, K. M. Arzayus, S. E. Walker, et al. "Automotive sources of carcinogenic polycyclic aromatic hydrocarbons associated with particulate matter in the Chesapeake Bay Region." Environ Sci Technol 2000; 34:4635-4640.

Dubowsky, S. D., L. A. Wallace and T. J. Buckley. "The contribution of traffic to indoor concentrations of polycyclic aromatic hydrocarbons." J Expo Anal Environ Epidemiol 1999; 9:312-321.

Dunbar, J. C., C. I. Lin, I. Vergucht, J. Wong and J. L. Duran. "Estimating the contributions of mobile sources of PAH to urban air using real-time PAH monitoring." Sci Total Environ 2001; 279:1-19.

Eder, E. "Intraindividual variations of DNA adduct levels in humans." Mutat Res 1999; 424:249-61.

El-Bayoumy, K. "Environmental carcinogens that may be involved in human breast cancer." Chem Res Toxi 1992; 5:585-590.

Fan, Z., D. Chen, P. Birla and R. M. Kamens. "Modeling of nitro-polycyclic aromatic hydrocarbon formation and decay in the atmosphere." Atmos Environ 1995; 27A:1139-1152.

Fertmann, R., I. Tesseraux, M. Schumann and H. Neus. "Evaluation of ambient air concentrations of polycyclic aromatic hydrocarbons in Germany from 1990 to 1998." J Expo Anal Environ Epidemiol 2002; 12:115-23.

Gammon, M. D., A. I. Neugut, R. M. Santella, S. L. Teitelbaum, J. A. Britton, M. B. Terry, et al. "The Long Island Breast Cancer Study Project: Description of a multi-institutional collaboration to identify environmental risk factors for breast cancer." Breast Cancer Res Treat 2002; 74:235-254.

Gammon, M. D., R. M. Santella, A. I. Neugut, S. M. Eng, S. L. Teitelbaum, A. Andrea Paykin, et al. "Environmental toxins and breast cancer on Long Island. I. polycyclic aromatic hydrocarbon (PAH)-DNA adducts." Cancer Epidemiol Biomarkers Prev 2002; 11.

Geolytics. CensusCD® Neighborhood Change Database (NCDB) 1970-2000. East Brunswick, NJ, Geolytics, Inc 2002.

Gunier, R. B., M. E. Harnly, P. Reynolds, A. Hertz and J. Von Behren. "Agricultural pesticide use in California: pesticide prioritization, use densities, and population distributions for a childhood cancer study." Environ Health Perspect 2001; 109:1071-8.

Hannigan, M., G. Cass, B. Penman, C. Crespi, A. Lafleur, J. Busby, WF, et al. "Bioassay-directed chemical analysis of Los Angeles airborne particulate matter using a human cell mutagenicity assay." Environ Sci Technol 1998; 32:3502-3514.

Harkov, R., A. Greenberg, F. Darack, J. M. Daisey and P. J. Lioy. "Summertime variations in polycyclic aromatic hydrocarbons at four sites in New Jersey." Environ Sci Technol 1984; 18:287-291.

Hu, P. S. and J. Young. Summary of Travel Trends 1990 Nationwide Personal Transportation Survey. Oak Ridge, Center for Transportation Analysis at ORNL 1992.

Huang, X.-D., D. Dixon and B. Greenberg. "Increased PAH toxicity following their photomodification in natural sunlight: impacts on the Duckweed Lemma Gibba L. G-3." Ecotox Environ Safety 1995; 32:194-200.

Johnson, W. B., F. W. Ludwig, W. F. Dabberdt and R. J. Allen. "An urban diffusion simulation model for carbon monoxide." JAPCA 1973; 23:490-498.

Levy, J. I., E. A. Houseman, J. D. Spengler, P. Loh and L. Ryan. "Fine particulate matter and polycyclic aromatic hydrocarbon concentration patterns in Roxbury, Massachusetts: a community-based GIS analysis." Environ Health Perspect 2001; 109:341-347.

Lewis-Michl, E. L., J. M. Melius, L. R. Kallenbach, C. L. Ju, T. O. Talbot, M. F. Orr, et al. "Breast cancer risk and residence near industry or traffic in Nassau and Suffolk, Long Island, New York." Arch Environ Health 1996; 51:255-265.

Li, D., W. Zhang, A. A. Sahin and W. N. Hittelman. "DNA adducts in normal tissue adjacent to breast cancer: a review." Cancer Detect Prev 1999; 23:454-62.

Lim, L. H., R. M. Harrison and S. Harrad. "The contribution of traffic to atmospheric concentrations of polycyclic aromatic hydrocarbons." Environ Sci Technol 1999; 33:3538-3542.

Lioy, P. L., J. M. Waldman, A. Greenberg, R. Harkov and C. Pietarinen. "The Total Human Environmental Exposure Study (THEES) to benzo(a)pyrene: comparison of the inhalation and food pathways." Arch Environ Health 1988; 43:304-12.

Long, C. M., H. H. Suh, P. J. Catalano and P. Koutrakis. "Using time- and size-resolved particulate data to quantify indoor penetration and deposition behavior." Environ Sci Technol 2001; 35:2089-2099.

Luhar, A. and B. Sawford. "An examination of existing shoreline fumiagtion models and formulation of an improved model." Atmos Environ 1996; 30:609-620.

Mitra, S. and B. Ray. "Patterns and sources of polycyclic aromatic hydrocarbons and their derivatives in indoor air." Atmospheric Environment 1995; 29:3345-3356.

Morgan, M. and M. Henrion. Uncertainty: A guide to Dealing with Uncertainy in Quantitative Risk and Policy Analysis. Cambridge, Cambridge University Press 1990.

Morris, J. J. and E. Seifter. "The role of aromatic hydrocarbons in the genesis of breast cancer." Med Hypotheses 1992; 38:177-184.

Naumova, Y., S. J. Eisenreich, B. J. Turpin, C. P. Weisel, M. T. Morandi, S. D. Colome, et al. "Polycyclic aromatic hydrocarbons in the indoor and outdoor air of three cities in the U.S." Environ Sci Technol 2002; 36:2552-2559.

NCDC. Hourly Surface Observations, Islip. Asheville, NC, National Climatic Data Center 1999a. 1999.

NCDC. Mixing Heights, Islip and Atlantic City for 1990 and 1993. Asheville, NC, National Climatic Data Center 1999b. 1999.

NCOPT. Hourly variation in traffic - 1977 - typical day at a typical location in Nassau County. Nassau County Unified Public Mobility Plan, Update. Mineola, Nassau County Office of Public Transportation 1978.

NCRP. Uncertainty in NCRP Screening Models Relating to Atmospheric Transport, Deposition and Uptake by Humans. Bethesda, National Council on Radiation Protection and Measurement 1993.

NDPW. Paper maps showing annual average daily traffic in Nassau County by year of measurement from the 1960s to the 1990s, Nassau Department of Public Works 1965-1995.

Nelli, J. P., A. D. Messina and J. A. Bullin. "Analysis and modeling of air quality at street intersections." JAPCA 1983; 33:760-764.

NJDOT. New Jersey total mileage and annual vehicle miles traveled by year, 1965-1990. Bureau of Transportation Data Development, New Jersey Department of Transportation Development. Trenton, Bureau of Transportation Data Development, New Jersey Department of Transportation Development 2002.

NYCDOT. 1997 New York City Screenline Traffic Flow. New York City, NYC Department of Transportation 1999a:111.

NYCDOT. Database of traffic counts in all five boroughs since 1992. New York City, New York City Department of Transportation, Freedom of Information Unit 1999b. 1999.

NYCDOT. New York City Bridge Traffic Volumes 1997. New York City, NYC Department of Transportation 1999c:191.

NYDOT. Traffic Volume Reports, 1970, 1980, 1990. Albany, New York Department of Transportaton 1991.

O'Brien, R. G. (1998). A tour of UnifyPow: a SAS module/macro for sample-size analysis. Proceedings of the 23rd SAS users group international conference, Cary, NC, SAS Institute 1998.

O'Brien, R. G. and K. E. Muller. Unified power analysis for t-tests through multivariate hypotheses. Applied analysis of variance in behavioral science. L. K. Edwards. New York, M. Dekker 1993:297-342.

Ohura, T., T. Amagai, T. Sugiyama, M. Fusaya and H. Matsushita. "Characteristics of particle matter and associated polycyclic aromatic hydrocarbons in indoor and outdoor air in two cities in Shizuoka, Japan." Atmospheric Environment 2004; 38:2045-2054.

Ortuzar, J. and L. Willumsen. Modelling Transport. Chichester, John Wiley & Sons 1995.

Ott, W., P. Switzer and N. Willits. "Carbon monoxide exposures inside an automobile traveling on an urban arterial highway." J Air Waste Manage Assoc 1994; 44:1010-8.

Palmer, L. Traffic growth rates for Eastern and Western Suffolk County. J. Beyea, Suffolk Department of Public Works 1999.

Raaschou-Nielsen, O., O. Hertel, E. Vignati, R. Berkowicz, S. S. Jensen, V. B. Larsen, et al. "An air pollution model for use in epidemiological studies: evaluation with measured levels of nitrogen dioxide and benzene." J Expo Anal Environ Epidemiol 2000; 10:4-14.

Ramsdell, J. V., Jr., C. A. Simonen and K. W. Burk. Regional Atmosphere Transport Code for Hanford Emission Tracking (RATCHET). Richland, WA, Battelle Pacific Northwest Laboratories 1994.

Reynolds, P., J. Von Behren, R. B. Gunier, D. E. Goldberg, A. Hertz and D. F. Smith. "Childhood cancer incidence rates and hazardous air pollutants in California: an exploratory analysis." Environ Health Perspect 2003; 111:663-8.

Sakai, R., H. C. Siegmann, H. Sato and A. S. Voorhees. "Particulate Matter and Particle-Attached Polycyclic Aromatic Hydrocarbons in the Indoor and Outdoor Air of Tokyo Measured with Personal Monitors." Environmental Research 2002; 89:66-71.

Sculley, R. "Vehicle emission rate analysis for carbon monoxide hot spot modeling." JAPCA 1989; 39:1334-1343.

SDPW. Paper maps showing annual average daily traffic in Suffolk County by year of measurement from the 1960s to the 1990s, Suffolk Department of Public Works 1965-1995.

Sheldon, L., A. Clayton, J. Keever, R. Perritt and D. Whitaker. PTEAM: Monitoring of Phthalates and PAHS in Indoor and Outdoor Air Samples in Riverside California. Sacramento, CA, California Environmental Protection Agency Air Resources Board (CEPAARB) 1992.

Sheu, H.-L., W.-J. Lee, K. P. Hwang, M.-C. Liow, C.-C. Wu and L.-T. Hsieh. "Dry deposition velocities of polycyclic aromatic hydrocarbons in the ambient air of traffic intersections." J Environ Sci Health 1996a; A31:2295-2311.

Sheu, H.-L., W.-J. Lee, J.-H. Tsai, Y.-C. Fan, C.-C. Su and H.-R. Chao. "Particle size distribution of polycyclic aromatic hydrocarbons in the ambient air of a traffic intersection." J Environ Sci Health 1996b; A31:1293-1316.

Sistla, G., P. Samson, M. Keenan and S. Rao. "A study of pollutant dispersion near highways." Atmos Environ 1979; 13:669-685.

Stellman, J. M., S. D. Stellman, T. Weber, C. Tomasallo, A. B. Stellman and R. Christian, Jr. "A geographic information system for characterizing exposure to Agent Orange and other herbicides in Vietnam." Environ Health Perspect 2003; 111:321-8.

USDOT. 1995 Nationwide Personal Transportation Survey Data Files, CDROM. Washington, DC, US Department of Transportation, Federal Highway Administration 1996. 1996.

USEPA. Summary Report for the Workshop on Monte Carlo Analysis. Washington, Risk Assessment Forum, Environmental Protection Agency 1996: 481.

USEPA. PCRAMMET User's Guide. Research Triangle Park, Environmental Protection Agency 1999:95.

USEPA. National Air Quality and Emissions Trends Report. 2003 Special Studies Edition, Chapter 2. Research Triangle Park, NC, Office of Air Quality and Standards, Air Quality Strategies and Standards Division, Report EPA 454/R-03-005 2003.

Venkataraman, C. and J. Raymond. "Estimating the lung deposition of particulate polycyclic aromatic hydrocarbons associated with multimodal urban aerosols." Inhal Toxicol 1998; 10:183-204.

Viegele, W. J. and J. H. Head. "Derivation of the gaussian plume model." JAPCA 1978; 28:1139-1141.

Viras, L. G., P. A. Siskos and E. Stephanou. "Determination of polycyclic aromatic hydrocarbons in athens atmosphere." Internat J Environ Anal Chem 1987; 28:71-85.

Ward, M. H., J. R. Nuckols, S. J. Weigel, S. K. Maxwell, K. P. Cantor and R. S. Miller. "Identifying populations potentially exposed to agricultural pesticides using remote sensing and a Geographic Information System." Environ Health Perspect 2000; 108:5-12.

Tables

Table 1. Data and methods used to estimate numbers of vehicles on the road by location, year, and operating condition.

	Warm engines		Cold engines	
Quantity of interest	Cruise	Acceleration/ deceleration	Cruise	Acceleration/ deceleration
Average vehicles per day (AADT).[a)]	Traffic counts from state and county traffic departments since the mid-1960s.[b)]	Same as cruise AADT	Trips per household per day times number of households per census block.[c)]	Same as cold cruise
AADT for missing years	Interpolation & extrapolation.[d)]	Same	Scaling by population and travel trends.[e)]	Same
Emission spatial restrictions	Within 80 km of Long Island boundaries.	Within 50 m of intersections.[f)]	Within 1-km of each census block centroid.	Same as cold cruise; also within 50 m of intersections. [f)]
% of vehicles slowed down or stopped at intersections		Simple function of traffic flow.[g)]		Same as for warm conditions

a) Annual average daily traffic counts (AADT) for locations without measurements were set equal to the nearest measured value on the same road for the same year, provided there was a measurement within 10 km. Sensitivity tests indicated that no significant difference resulted in the exposure distribution down to a 2-km distance cutoff. On Long Island, the median distance to the nearest traffic count on major roads was 0.6 km, with a mean distance of 1.4 km.

b) Paper maps with numerical counts marked next to roads were available for Nassau and Suffolk counties from 1964-1991 (SDPW 1965-1995), (NDPW 1965-1995). The paper maps were compared section by section with GIS maps displayed on a laptop screen. Corresponding locations of a count along a road were estimated by eye and marked with a cursor click, which caused the latitude and longitude to be stored in a table and a screen form to pop up allowing entry of the numerical count value read from the map. Programming was done in MapBasic. Once the traffic count value was entered, it was backchecked against the original value on the paper map. Data from the 1990s for the entire 22-county, NY Metropolitan area (NYMSA) were purchased from (BLR 1997). Recent traffic count data for Queens County were obtained from the City of New York (NYCDOT 1999b). Historical traffic growth rates for the outer counties came from data on state roads (CONNDOT 1991), (NYDOT 1991), (NJDOT 2002). Historical data for New York City were obtained by backwards scaling using measured traffic flows on New York City bridges (NYCDOT 1999a; NYCDOT 1999c) and between boroughs.

c) Trips were allocated to major roads within one km of the centroid of the census block, using a standard inverse distance square law weighting (Ortuzar and Willumsen 1995). Trips per household per day for Long Island and the NYMSA were extracted from the database available through (USDOT 1996). Household numbers came from (Census_Bureau 1993)

d) If multiple-year data were available at a point within a ten-year period, using the method in the first footnote, then logarithmic interpolation or extrapolation was used to fill in traffic count values for a missing year. Otherwise, countywide growth rates were used, with the exception of Suffolk County where separate rates were used for the eastern and western parts of the county (Palmer 1999). County growth rates were obtained by analyzing the historical values of traffic counts on state roads. In this way, AADT were assigned to 99% of the street locations.

e) Numbers of households were scaled by census tract figures from 1960-2000 (Geolytics 2002). Trips per household per day were scaled by historical values of trips per person per day (Hu and Young 1992).

f) The intersection distance was broken down into 3 sections, 50, 25, and 12.5 meters, allowing a graded rate of emissions. In the default model, the relative emissions were taken in the proportion of 1:2:3.

g) For unequal traffic counts on intersecting roads, we made the relative stopping fraction a function of the relative traffic counts. Even for very high volume roads, we assumed that 10% of the traffic was exiting or entering, and therefore accelerating or decelerating. This approach is a simplification of more elaborate models used to predict peak CO exposures at intersections (Nelli, Messina et al. 1983).

Table 2. Summary of BaP vehicle emission modeling parameters abstracted from the literature (default values)

Ratio of cold cycle to warm cycle	8 ± factor of 4
Distance traveled before engine warms up	1-km
Ratio of winter to summer emissions	< 2
Ratio of emissions at intersections to cruise emissions	10
Duration of acceleration/deceleration period at intersections	Undetermined

Table 3. Default meteorological dispersion model

Quantity of interest	Data source	Comment	Options for sensitivity analysis
< 100 m of road	Chock dispersion model (Chock 1978).[a]	Models merged at 100 m.[b]	
> 100 m of road	Gaussian plume model (Viegele and Head 1978).	Briggs dispersion parameters used. (Catalano, Turner et al. 1987).[b]	Rural dispersion parameters
Meteorological data for wind speed and direction.	1993 values from Brookhaven National Labs meterological station (BNL)		1990 BNL values; 1990 and 1993 data from MacArthur airport(NCDC 1999a).
Mixing layer	1993 values from (NCDC 1999b).[c]		1990 values.
Rain washout	Hourly mm of rainfall (BNL).	Took standard function of rainfall rate (Ramsdell, Simonen et al. 1994).	Scale factor
Photo-decay	Hourly pyranometer readings from (BNL).	Proportional to the amount of sunlight, adjusted to give an average 6-hour decay rate.[d]	Scale factor. 1990 pyranometer readings.
Deposition velocity	(NCRP 1993)	Default value = 0.003 m/s	Scale factor

a) Computationally, the Chock model was converted to a gaussian plume formulation by fitting the Chock predictions to spatially dependent, gaussian plume dispersion parameters applied to puffs emitted along the roadway.

b) At 100 meters from the road, the puffs were allowed to continue to expand using standard dispersion parameters, whose values for the default model were taken from 1993 Brookhaven Laboratory measurements (BNL). Values measured at Brookhaven in 1990 and values derived using PCRammet from data collected at MacArthur airport in 1990 and 1993 were used for sensitivity studies (USEPA 1999), (NCDC 1999a).

c) Raw data converted to hourly mixing heights using the program, PCRammet (USEPA 1999).

d) The default value was chosen to produce a mean BaP lifetime during daylight of 6 hours, a value within the wide range given in the literature (Huang, Dixon et al. 1995), (Fan, Chen et al. 1995).

Table 4. Statistics of residential address information (Nassau and Suffolk Counties) recorded at interview

	Number of locations	Percentage
Total residential locations	8321	100
Locations without streets	300	3.6
Locations with streets, but no numbers	1715	21
Locations with street numbers	6306	76
Locations geocoded to street level	5501	66 (87[a])

a) Percentage of locations with street numbers that geocoded successfully to the street level.

Table 5. Sensitivity of test statistic to variations in model.

Test statistic[a] category	Residential Exposure
Baseline (Standard error of difference from sample size limitations) [b]	5.5%
Different deposition velocities	1.9%
Meteorologic data for a different year	1.8%
Include washout from precipitation	0.2%
Exclude contribution from distant counties	0.9%
Variable penetration rate of fine particles[c]	1.1%
No Intersections	1.4%
Different PAH source term	0.8%
Root-mean-square combination of model uncertainties[d]	3.5%

a) Let M1 and M2 be the exposure means for a split sample. The test statistic is simply the difference in means divided by the average = (M1-M2)/((M1+M2)/2). This quantity varies for different splits of the study population and for different parameter choices in the model. The root-mean-square average of this quantity for many splits of the study sample is presented in the table along side the parameter varied

b) Each split sample contains approximately 1500 women. The percentage differences presented in the table are rms averages over the results of 40 different splits, all of which were selected randomly.

c) This is a variability, not an uncertainty.

d) Baseline not included.

Table 6. Number of women in study population available for analysis, regardless of when they moved to Long Island.

Exposure category and constraint on percentage of missing address information[a]	Number of women	Women excluded (< 5 yrs of exposure during period)	Net total
Cumulative 1960-1990[b]			
Address info for 100% of period	1452	89	1363
Address info for 80% or more of period	1932	96	1836
Address info for 75% or more of period	2026	96	1930
Address info for 50% or more of period	2283	100	2183
Any address info for period	2534	102	2432
Dose under age 20[c]			
Address info for 100% of period	371	100	271
Address info for 80% or more of period	470	105	365
Address info for 75% or more of period	484	108	376
Address info for 50% or more of period	530	119	411
Any address info for period	567	129	438
Dose before Age first birth[c]			
Address info for 100% of period	328	192	136
Address info for 80% or more of period	482	204	278
Address info for 75% or more of period	514	207	307
Address info for 50% or more of period	628	226	402
Any address info for period	735	240	495

a) A complete address is one for which street-level geocoding occurs.
b) If a woman arrives after 1960, the start period is her arrival date.
c) If a woman was born before 1960, the start period is either 1960 or her arrival date, which ever is most recent

Table 7. Number of participants who arrived on Long Island before the start of various time periods, as a function of completeness of their address information.

	Time period					
	1960-1990	1965-1990	1970-1990	1975-1990	1980-1990	1985-1990
Completeness of address information[a)]						
100% of period	569	831	1122	1476	1749	1976
80% or more of period	832	1122	1402	1709	1920	2117
75% or more of period	879	1164	1454	1757	1970	2142
50% or more of period	1016	1347	1628	1921	2102	2248
Any address info for period	1160	1506	1787	2066	2220	2309

a) Percentage of time for which a street-level latitude and longitude is available. A complete address is one for which street-level geocoding occurs.

(Figures are in a separate PDF file , 4 mb in size.)

Figure Captions.

Figure 1. LIBCSP study area showing the major roads within an 80-km distance of Long Island from which vehicle emissions are tracked in this study. Study participants were drawn from the shaded area, which is 150-km in length. The straight line crossing Long Island defines the location of predicted air concentrations shown in Figure 7.

Figure 2. Traffic-count measurement density in a 25-km wide, section of study area. Each symbol represents a measurement location.

Figure 3. Relative vehicle emissions for warm-engine and cold-engine conditions by hour of day. Warm-engine data collected at a typical location in Nassau County in 1977. Cold-engine data averaged over Nassau and Suffolk, based on 1995 traffic diaries. See text.

Figure 4. Historical tailpipe emissions of PAH reconstructed for this study

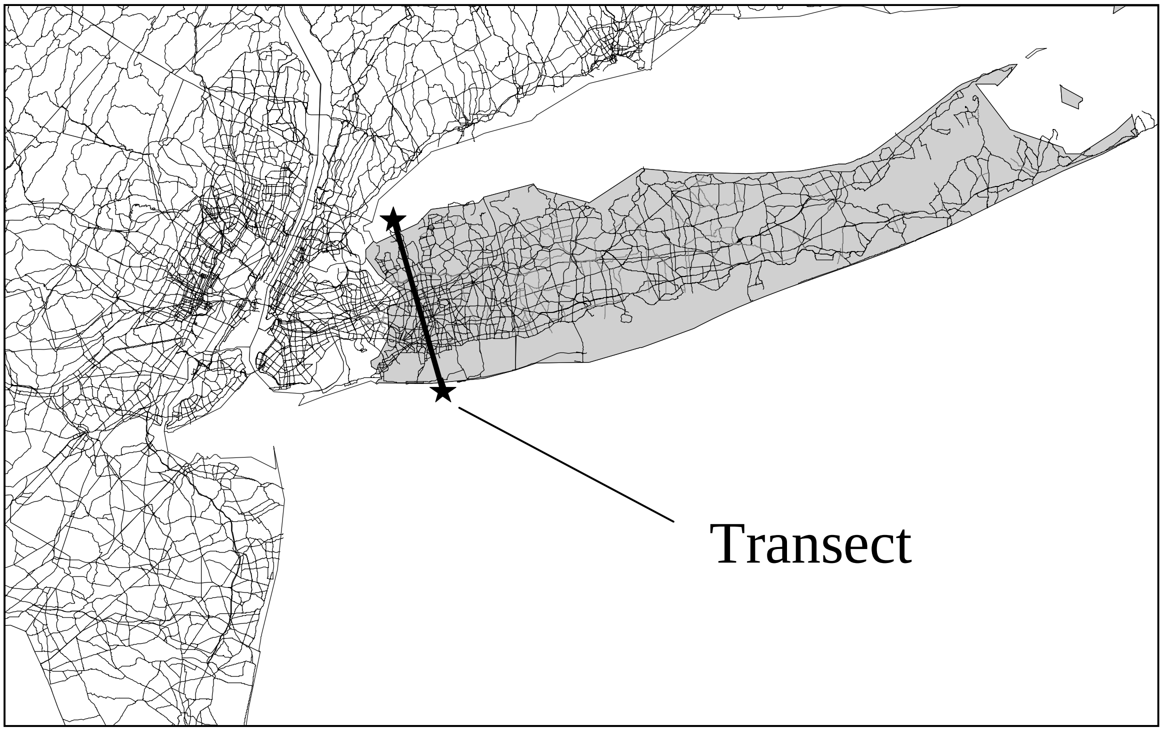
Figure 5. Geocoding success to high accuracy by year (5 year average).

Figure 6. Segments of roads within 100 meters of major intersections where PAH emissions are increased. Shown as cross-like structures, darkened for visual emphasis, within a 4-km by 6-km wide map selection from Nassau County, Long Island, New York.

Figure 7. Predicted PAH exposures along a transect across Long Island. See Figure 1 for the location of the transect.

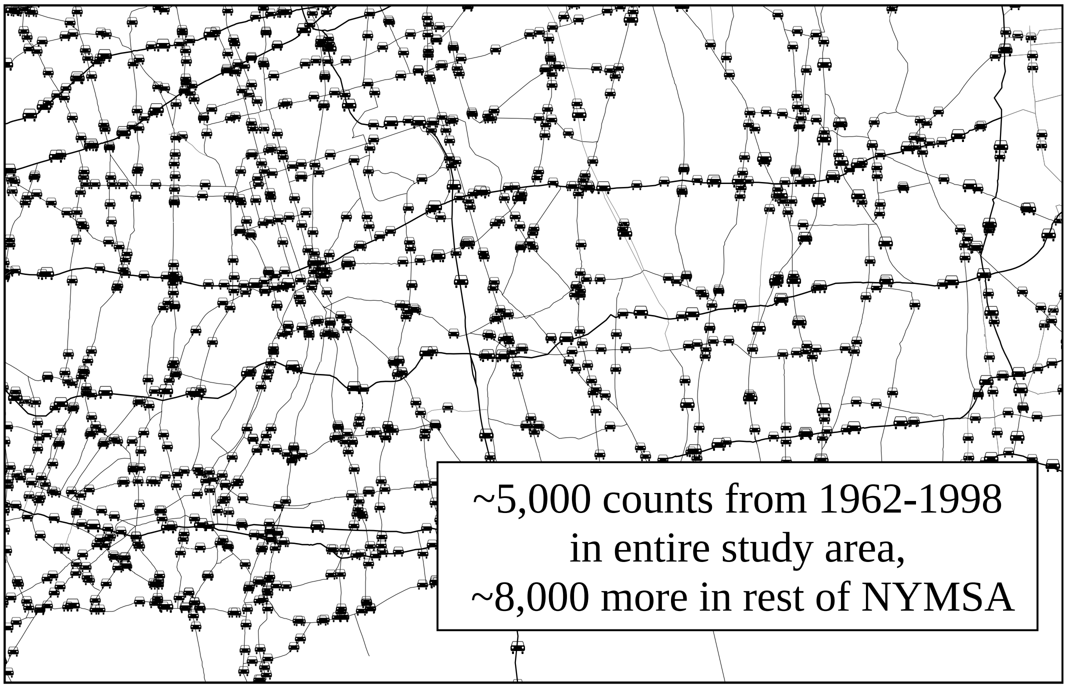
Figure 8. Yearly exposure to traffic PAH for hypothetical subject. Relative units.

Figure 9. Exposure opportunity index by year of arrival on Long Island. Data shown for 2534 women with geocodable addresses, both cases and controls, who were participants in the Long Island Breast Cancer Study Project.Figure 1



LIBCSP study area showing the major roads within an 80-km distance of Long Island from which vehicle emissions are tracked in this study. Study participants were drawn from the shaded area, which is 150-km in length. The straight line crossing Long Island defines the location of predicted air concentrations shown in Figure 7.

Figure 2


~5,000 counts from 1962-1998
in entire study area,
~8,000 more in rest of NYMSA

Traffic-count measurement density in a 25-km wide, section of study area. Each symbol represents a measurement location.

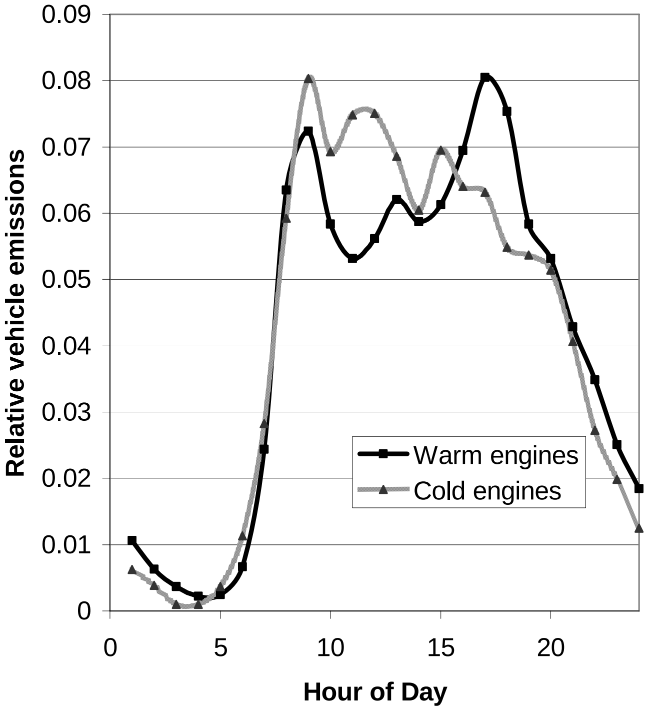
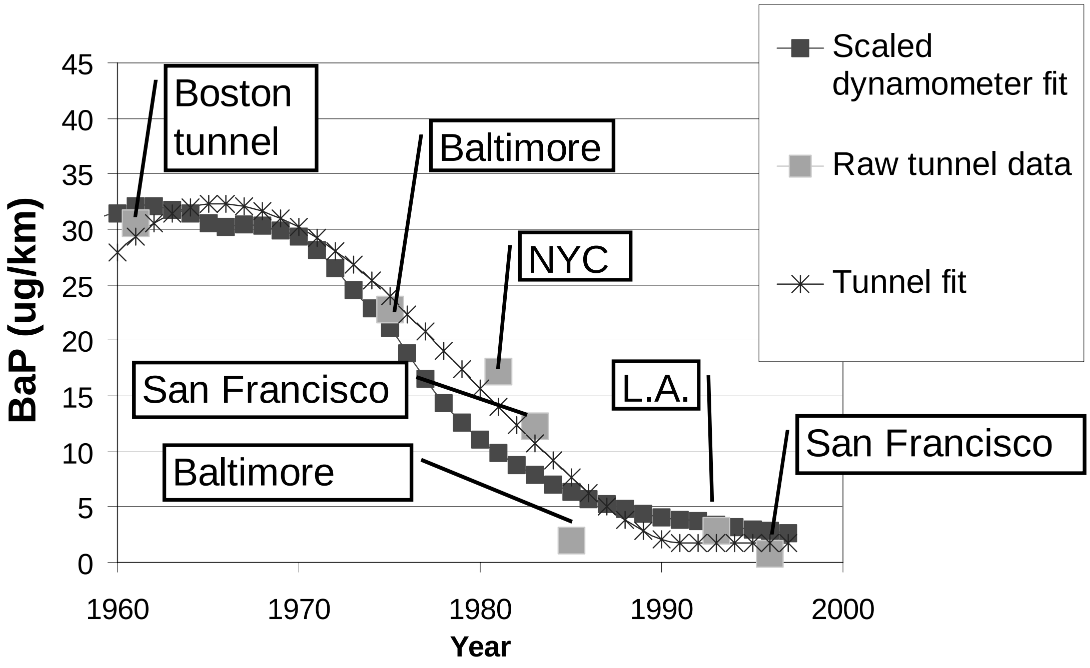


Figure 3

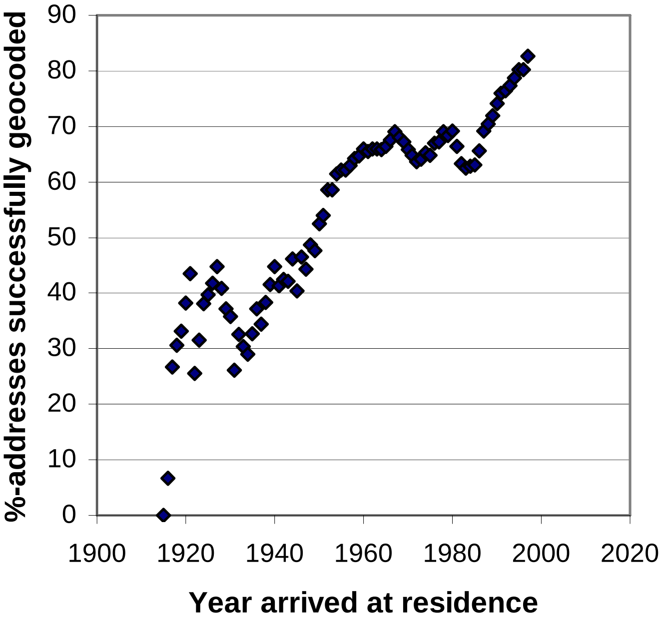
Relative vehicle emissions for warm-engine and cold-engine conditions by hour of day. Warm-engine data collected at a typical location in Nassau County in 1977. Cold-engine data averaged over Nassau and Suffolk, based on 1995 traffic diaries. See text.

Figure 4.



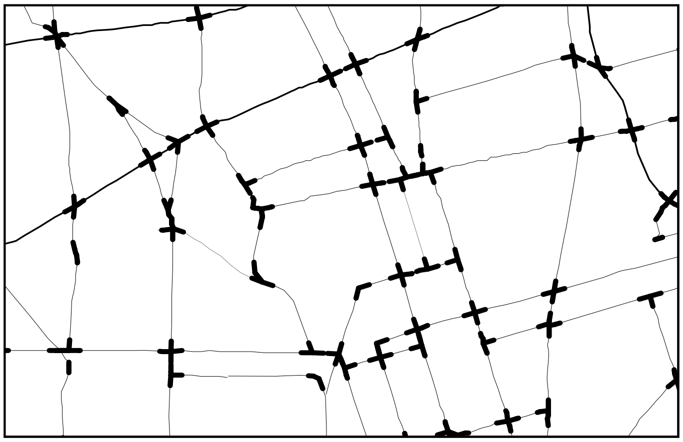
. Historical tailpipe emissions of PAH reconstructed for this study

Figure 5



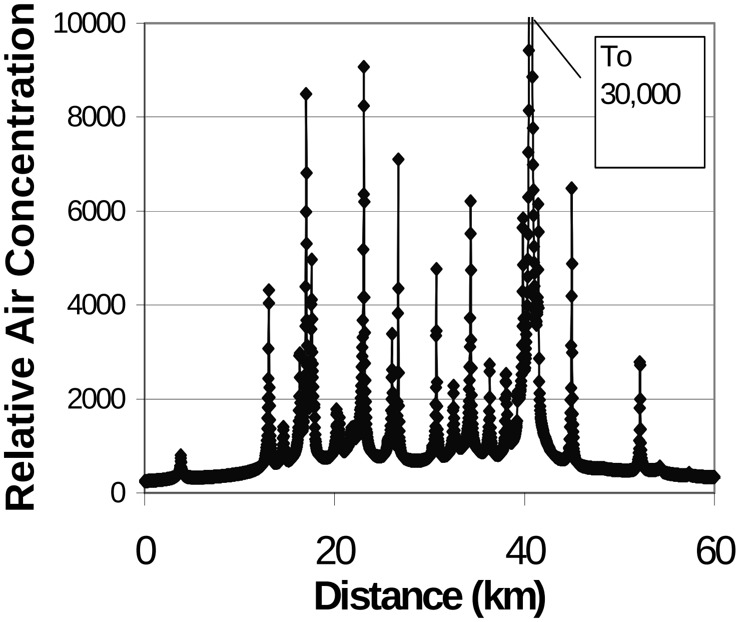
Geocoding success to high accuracy by year (5 year average).

Figure 6.



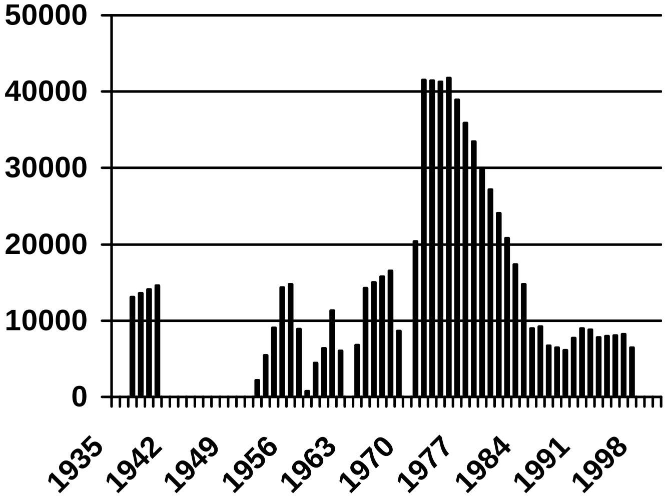
Segments of roads within 100 meters of major intersections where PAH emissions are increased. Shown as cross-like structures, darkened for visual emphasis, within a 4-km by 6-km wide map selection from Nassau County, Long Island, New York

Figure 7.



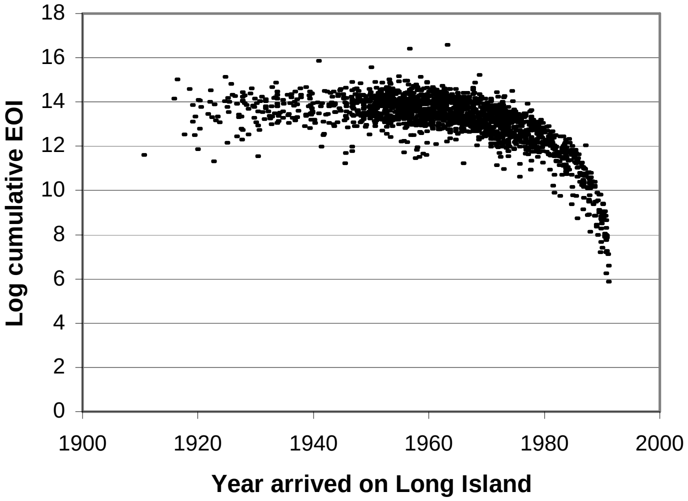
Predicted PAH exposures along a transect across Long Island. See Figure 1 for the location of the transect.

Figure 8.



Yearly exposure to traffic PAH for hypothetical subject. Relative units

Figure 9.



Exposure opportunity index by year of arrival on Long Island. Data shown for 2534 women with geocodable addresses, both cases and controls, who were participants in the Long Island Breast Cancer Study Project.