National Human Exposure Assessment Survey (NHEXAS)

Maryland Study

Quality Systems and Implementation Plan for Human Exposure Assessment

Emory University
Atlanta, GA 30322

Cooperative Agreement CR 822038

Standard Operating Procedure **HSPH/NHEXAS/QSIP**

Title: Quality Systems and Implementation Plan for Pilot Studies for NHEXAS

Source: Harvard University/Johns Hopkins University

U.S. Environmental Protection Agency
Office of Research and Development
Human Exposure & Atmospheric Sciences Division
Human Exposure Research Branch

Notice: *The U.S. Environmental Protection Agency (EPA), through its Office of Research and Development (ORD), partially funded and collaborated in the research described here. This protocol is part of the Quality Systems Implementation Plan (QSIP) that was reviewed by the EPA and approved for use in this demonstration/scoping study. Mention of trade names or commercial products does not constitute endorsement or recommendation by EPA for use.*

QUALITY SYSTEMS AND IMPLEMENTATION PLAN (QSIP)

for

Pilot Studies for NHEXAS

by

P. Barry Ryan
Harvard School of Public Health
Boston, Massachusetts 02115

TITLE AND APPROVAL SHEET

QUALITY SYSTEMS AND IMPLEMENTATION PLAN (QSIP)

for

Pilot Studies for NHEXAS

implemented by
HSPH/Emory/JHU Consortium

composed of

Rollins School of Public Health of Emory University
Atlanta, GA 30322
Harvard School of Public Health
Boston, MA 02215

Johns Hopkins University
Baltimore, MD

_____
P. Barry Ryan, PhD:
Professor (Emory) and
Adjunct Associate Professor of
Environmental Health (Harvard)
Principal Investigator (PI)

_____
Thomas A. Burke, PhD:
Associate Professor of Environmental Health

_____
Robert Weker
Quality Assurance Officer

ENVIRONMENTAL PROTECTION AGENCY
oversight by

_____
Karen Hammerstrom
EPA Principal Collaborator

_____
David Kleffman
NHEXAS Project Officer

TABLE OF CONTENTS

1.0 PROJECT PLANNING AND ORGANIZATION

1.1 Project Scope and Work Objectives

The National Human Exposure Assessment Survey (NHEXAS) is an EPA program with two principal objectives: (1) to document the status and trends in the national distribution of pollutant exposure through statistically representative sampling of the U.S. population, and (2) to determine the primary causes of these exposures by studying household, local, and regional characteristics. The investigation described in this document is designed to be exploratory in nature with principal aims directed toward development and field evaluations of survey designs and associated sampling and analytical methodologies. In this regard, the investigations fit under what has been called the "Phase I" studies for NHEXAS. The utility of the results obtained, however, is not limited to the NHEXAS program. The particular results of this investigation will improve understanding of multimedia exposures, factors influencing such exposures, and the temporal variability associated with such exposures. This information alone would make the study worth undertaking. In conjunction with the NHEXAS program, the improved understanding of the temporal component of the variability will afford improvement in the survey design for the full-scale, national NHEXAS investigation known as Phase II.

The collaborating organizations, Harvard School of Public Health (HSPH), Rollins School of Public Health of Emory University, Johns Hopkins University (JHU), Westat Inc. (Westat), and Southwest Research Institute (SwRI), will compare alternative sampling and analysis protocols to investigate their ability to afford accurate estimation of long-term multimedia exposure to pollutants from three general classes: metals, pesticides, and polycyclic aromatic hydrocarbons (PAHs). Specifically, the investigation will look at temporal trends in indoor air, outdoor air, personal air, water, soil, house dust, dermal wipes, and food media as well as biological measures from blood and urine. Table 1 shows the analytes to be studied.

Table 1 -- Analytes for HSPH NHEXAS Phase I Study

Category	Analyte	CAS number	Category	Analyte	CAS number
Metals	Arsenic	7440-38-2	Pesticides	Chlordane	57-74-9
	Cadmium	7440-43-9		Chlorpyrifos	2921-88-2
	Chromium	7440-47-3		4,4'-DDD	72548
	Lead	7439-92-1		4,4'-DDE	72559
PAHs	Anthracene	120-12-7		4,4'-DDT	789-02-6
	Benzo(a)pyrene	50-32-8		Dieldrin	60-57-1
	Chrysene	218-01-9		Heptachlor	76-44-8
	Phenanthrene	85-01-8		Malathion	121-75-5

Table 2 shows the media that will be studied, the analytes for which each medium will be analyzed, and what organization will be responsible for analysis. (In some cases an agency will send samples to contract laboratories for analysis.)

Table 2 -- Media, Analytes, and Analyzing Organization

Medium	Metals (4)	Pesticides (8)	PAHs (4)	Total
Indoor Air	4 Emory	8 SwRI	4 SwRI	16
Outdoor Air	4 Emory			4
Personal Air	4 Emory			4
House Dust	4 Emory	8 SwRI	4 SwRI	16
Soil	4 Emory	8 SwRI	4 SwRI	16
Dermal Wipe	4 Emory	8 SwRI		12
Water	4 EPA	8 EPA	4 EPA	16
Food	4 FDA	8 FDA	4 FDA	16
Urine	3 CDC	2 CDC		5
Blood	2 CDC	6 CDC	4 CDC	12
number of analyses	37	56	24	117

The series of metals and compounds chosen span a range of likely concentrations and source characteristics. While not all-inclusive, they were chosen to include substances likely to have differing characteristics with respect to distribution of exposure in the population and over time.

Households, each including a target individual, will be asked to participate for eight seven-day Cycles over the course of a year. Enough households will be chosen to make it probable that at least 50 will complete at least six Cycles. (See Section 2.1.1.)

The primary focus of the Harvard/Emory/Johns Hopkins/Westat investigation is the evaluation of the temporal variability in pollutant exposures. The primary hypothesis to be tested can be concisely stated as:

- Pollutant exposures estimated from short-duration measurements are equivalent to actual exposures experienced over much longer durations (e.g., one year).

Implicit in this hypothesis is a working definition of pollutant exposures. In this document, we use the definition that exposure requires the simultaneous presence of a pollutant concentration and a human receptor (See Sexton and Ryan, 1988). Exposure is related to dose, delivery of the pollutant to the target organ, but is not identical. Exposure can occur without the individual receiving a dose to a target organ. The study has somewhat conflicting needs. One major component of this investigation is to improve understanding of the ability of a short-term measurement of exposure to estimate the long-term exposure experienced by an individual. Under this scenario, each individual measure of exposure, including each medium evaluated and each pollutant measured, can be considered separate. All have been used as short-term measures of exposure and thus are eligible for investigation in this study. In effect, then, as shown in Table 2, the study calls for 117 different exposure measures using 10 media and 16 pollutants. The appropriate values for evaluation under this scenario are the

distributional parameters, including means, standard deviations, skewness, kurtosis, and non-parametric statistics such as the median, 75th percentile, etc., for each of the 117 different exposure measures. These would have to be evaluated for the full population under investigation for each monitoring period. Certain statistical problems would arise from this many comparisons. For example, multiple comparisons can lead to associations that are statistically significant but still occur by chance. If one adopts an acceptance criterion for Type I error of 5%, then in 117 analyses, one would expect about 6 to give statistically different results even if the null hypothesis were true. This must be accounted for in any statistical analysis through adjustment of the criterion for allowed Type I error.

Alternatively, one may choose to use the exposure experienced for each of the compounds, aggregated over media: a total of 16 different exposure measures. We will refer to this as "total exposure." Total exposure requires information to be gathered on the relative impact (or contribution) of each medium on exposure. This information can be collected through questionnaire data and through activity diaries. The appropriate values for evaluation under this scenario would be the distributional parameters for total exposure for the population under investigation, again for each monitoring period, and the "averaged" exposure distribution representing the entire year's exposure. The advantage of this approach is that a true exposure estimate is used to assess the efficacy of using short-term measurements for long-term estimation. The main drawback to this approach is the error introduced by using models to relate the measurement to total exposure. No validated models exist that relate environmental and biological measurements to true exposures. The models are uneven in quality and are likely to give errors of unknown structure (e.g., differing biases, heterogeneity in errors among groups, etc.). This calls into question statistical analyses based upon normally-distributed, unbiased errors.

Another approach to this is to use biological markers, i.e., the blood and urine measurements, as indicators that exposure has taken place. One could then construct models relating environmental, questionnaire, activity, and any other measures to these, thereby establishing them as the fundamental exposure outcome variables. The chief advantage of this approach is the ease of analysis and interpretability of results. General linear modeling provides methods for rapid determination of the effects of various environmental media on the biological sample. Additionally, categorization of respondents into classes is possible using models based upon questionnaire data. The downside to such an analysis is the assumption of equivalence between the biological measures (and the concomitant inter-individual variability in, for example, metabolic rates) and true exposure. Additionally, these biological measures are more appropriately measures of does, not exposure.

These analysis procedures supply important data, not only for the NHEXAS investigations, but also for any series of studies operating in a single medium, multiple media, and for multiple pollutants. Our primary evaluation will be through the total exposure variable. Using techniques developed at Harvard (MacIntosh et al. 1994a; MacIntosh et al. 1994b), we will estimate total exposure from all media for each individual, during each monitoring period. We will then evaluate the parameters of these distributions. On the other hand, data will also be available to address hypotheses associated with each individual medium and pollutant combination.

Although our main focus will be assessment of the relationship between short-term measurements of exposure and long-term exposure estimates, other hypotheses can also be investigated. One secondary hypothesis to be tested is:

- Questionnaire data are sufficient to afford prediction of exposures experienced by a study population.

Using the data collected in the detailed questionnaire, it may be possible to categorize individuals into

differing classes of exposure. We propose to investigate this hypothesis with our study population.

An additional secondary hypothesis addresses exposures experienced through the food medium. The standard method of assessing the food-medium exposure is through analyses of duplicate diet samples. This is an expensive process. Alternative strategies requiring less expense and less burden on the respondent offer attractive substitutes. In particular:

- A food use diary, coupled with a mini-market basket data collection procedure, is equivalent to a duplicate diet procedure to estimate exposures from this medium.

Our detailed study designed to assess these hypotheses is given below. The data collection effort outlined will result in the accumulation of data related to time variability in exposures that have not been obtained prior to this study. Thus the value of this investigation as an exploratory study should not be underestimated.

If the data obtained are not sufficient to ensure adequate testing of any hypothesis, they still remain valuable in terms of the final scientific needs of the NHEXAS pilots: to afford better designs for the full, Phase II NHEXAS investigations. Further, such data will supply new and important information to improve the understanding of the factors influencing total exposure.

1.2 Project Description

1.2.1 Overview of the Harvard/Emory/Johns Hopkins Study

The overall purpose of the NHEXAS studies is to gather information on human exposure to various classes of pollutants. Two of the sub-investigations are to draw samples in a population-based sense and gather such exposure data. The purpose of the Harvard/Johns Hopkins study is to investigate the relationship between short-term exposure measurements of various kinds and a long-term exposure estimate. To accomplish this task, our design invokes several components:

- Measurements of selected contaminants in indoor air, outdoor air, personal air, house dust, soil, dermal wipes, water, and food, and in two biological media: blood and urine.

- Administration of questionnaires eliciting information on the respondent's activities and on possible sources of target chemicals in his or her environment.

- Measurements will take place repeatedly over a 12-15 month period. These "Cycles" of measurements will use short-term protocols of one-day to one-week duration, spaced approximately six weeks apart. A total of eight such Cycles will be completed.

- Certain of the measurements will measure exposure concentrations directly. These include personal air measurements and duplicate diet measurements
- The other measurements (indoor air, outdoor air, house dust, soil, dermal wipes, water, urine, blood) must be combined through modeling, questionnaire response, data from other sources, and assumptions to produce estimates of exposure and dose. These indirect measurements are the bulk of the data collected.

Given this overview of the study design, we can now describe the details of the investigation. The Harvard/Johns Hopkins NHEXAS Pilot Study is designed to improve our understanding of the relationship between short-term exposure measurements and long-term exposure estimates. We define

exposure measurements as physical measurements of exposure-related concentrations. For air monitoring, this may be done by affixing some type of monitor to an individual and measuring concentrations of environmental variables over a period of time. For other media, instrumentation is not readily available, but, in theory, similar apparatus can be envisioned that acts in an equivalent fashion. It is also theoretically possible to take long-term measurements of exposure by outfitting an individual with a monitor for extended periods of time. Generally this is not feasible. Monitors are too bulky and inconvenient or may not exist for a pollutant or medium of interest. Exposure estimates, on the other hand, may use a combination of measurement techniques and modeling to effect an approximation of the values that would have been obtained had a true measurement been made. This is an important distinction that is often ignored.

Historically, exposure assessment field studies have gathered data on exposures experienced by groups of individuals over relatively short time spans, typically one week. The reasons for this are primarily logistic; participation in such a study is very burdensome to the individual and expensive to the funding agency. Further, from the scientific point of view, the quality of the information gathered is uncertain. Few studies have attempted to relate the results of such short-term investigations with measured, long-term exposures. Often the assumption is made *ad hoc* that such measurements over short durations can be directly extrapolated, that is, that they are equivalent.

As part of the TEAM investigations, Wallace et al. (1984) measured concentrations of VOCs in air, water, food, breath, dust, hair, blood, and urine on a small group of individuals; but most of the small number of investigations attempting to establish the relationship between short-term and long-term exposures have been restricted to a single medium, usually air (Spengler et al. 1981, Quackenboss et al. 1983, Spengler et al. 1994, Needleman et al. 1985). These studies focus on equipping respondents with some form of personal monitor that gathers data and allowing them to proceed through their daily activities. The typical study of this type repeats this monitoring protocol at regular intervals throughout a longer time period, typically a year. These studies provide insight into the relationship in question, although they are restricted to a single medium. While not gathering true exposure data, using our definition, the exposure estimates made in this fashion are still superior to those made based on a single-measurement protocol.

It is upon these studies that we build our design. Our investigation will make use of this repeated-measures design to gather data on the variability in short-term measurements over time. Further, the aggregation of the series of short-term measurements will afford an estimate of long-term exposures. The research outlined in this document is new in that all media will be considered for a suite of pollutants (see Section 1.1). The exposure estimates made, therefore, will include information about any potential correlations in environmental exposures among various pollutants.

1.2.2 Issues in Statistical Survey Design

We propose an investigation in which data will be gathered from a group of individuals repeatedly over the course of a one-year period. We will discuss below the reasoning behind the actual number of repeated measurements. In this section, we discuss the issues associated with statistical representativeness and generalizability.

In addressing a survey design, it is important to understand the statistical character of the sample being taken. The measurement protocol we are proposing is quite complex; it involves monitoring for a series of pollutants in various media over repeated sampling Cycles. In order to assess the appropriate sample size one must have information on several important parameters. Most significant among these from a design point of view are the expected variance in the target parameter within the distribution and the expected biases associated with the sampling design.

In our study the expected variance in target parameters is, itself, a complicated issue. We are

attempting to gather information about temporal variability in total exposures experienced through several media for several pollutants. There is no single target parameter upon which our design rests; total exposure is dependent on numerous pathways of exposure with variances in pollutant concentration, contact rate, and uptake influencing our design. Further, the variance in many, if not most, of these components with time is unknown. One could make *ad hoc* assumptions about the variance, perform some simulations, and assess the appropriate sample size, but such an exercise would give results completely dependent upon the assumptions made.

Alternatively, one may choose to base the design on an exposure with well-known characteristics. For example, one may choose soil lead contamination variability as a well-established parameter with known spatial and temporal variability. One would then design an investigation that, notwithstanding any biases in the selected sample, would result in a study that, within sampling error, would be generalizable to the population. Unfortunately, it would only be so with respect to this particular route of exposure to this pollutant. Other routes may be more or less adequately covered. Without a knowledge of the spatial and temporal variability of the other parameters, little could be said about their generalizability.

The expected biases in the sample population also exacerbate the design problems. The protocol outlined is very burdensome. The respondent is expected to perform numerous activities: filling out questionnaires, carrying personal monitors, supplying biological samples, keeping track of activities, and preparing duplicate meals. Further, the amount of equipment brought into the home will increase the "nuisance" factor associated with this study. One may expect that only individuals highly committed to scientific endeavors or those in need of the monetary remuneration offered by this study will participate. This serious sampling bias we expect impacts directly on the survey design. However, we have no *a priori* reason to expect the sample to be biased with respect to variability in short-term estimates of long-term exposures, or to the ability of certain sampling techniques to predict such exposures.

The final constraint on this design is pragmatic. A repeated-measures design necessarily requires multiple visits to the same individual. Each visit increases the number of samples to be analyzed. In surveys such as this, analytical chemistry costs often account for two-thirds or more (Pellizzari 1992) of the total cost. Further, the large number of data to be managed due to multiple sample types and multiple sampling periods sets feasibility limits on the investigation.

The longitudinal design of this study is a new tack in environmental exposure assessment and the data to be collected will provide information not yet available for many media. Consequently, very few data exist that can be used as inputs to a statistically based study design. Nevertheless, as described in the following sections, we have combined fundamental statistical principles in a novel manner to estimate the number of samples required to address the hypotheses presented previously. Our study design balances the need to collect a sufficient number of samples to yield new insights about temporal trends of multimedia pollutant exposures against the analytical chemistry costs of quantifying pollutant concentrations in environmental samples.

1.2.3 Potential Stratification Variables

In developing our survey design, we began to consider early on whether stratification of the sample would be appropriate. According to statistical survey design literature (see Kish, *Survey Sampling,* John Wiley & Sons, New York 1965), stratification of a sample is typically done for one of three reasons (quoted from Kish):

1. Stratification may be used to decrease the variances of sample estimates.
2. Strata may be formed to employ different methods and procedures within them.

3. Strata may be established because the sub-populations within them are also designated as domains of the study.

In our investigation, reason 2 is not appropriate in that all strata, no matter what criteria may be used, will be subject to the same protocol. We are left, then, to explore reasons 1 and 3 for potential stratification variables. Reason 1 approaches the essence of the use of stratification. To quote Kish:

> "... The variance [in sample means] is decreased to the degree that the stratum means diverge and that homogeneity exists within strata...."

If the variance within the individual strata is reduced sufficiently, an improved efficiency can be gained for population estimates through stratification. Further, estimates of stratum-specific parameters can also be made, although the reduced sample size begs caution in using such estimates. Thus, Reason 3 may also be operative if we are particularly interested in one or more strata of individuals with respect to exposure.

We will focus our discussion on stratification to decrease variance in sample estimates but note the utility of sampling specific sub-groups as well. Several potential stratification variables present themselves for consideration. Among those considered, three have the strongest arguments in their favor. These are age, socioeconomic status (SES), and geography. We now discuss the merits and deficiencies of each.

Age

Age-based stratification for exposure offers an attractive design for several reasons. It is commonly accepted that exposures to various age groups differ substantially. For example, infants and toddlers exhibit mouthing behaviors, which may result in elevated exposure to soil-borne pollutants through ingestion. Young children may be subject to more dermal contact than adults due to the larger amount of time spent in outdoor activities. Adults may be expected to have increased exposures due to occupational, transit-related, and perhaps other exposure modes.

On the other hand, exposure from certain media are unlikely to be substantially different. For example, airborne pollutants are likely to be distributed without significant stratification among ages. Similarly, pollutants in drinking water may not exhibit significant differences among age groups.

One can envision several strata: infants (< 6 months); toddlers (6 months - 5 years); young children (ages 6-12); adolescents (13-18); and adults (>18). Because of the complexity of our investigation, the need for personal monitoring, the collection of blood and urine biological samples, and the need for a basic understanding of the motivation behind our work, we plan on allowing participation only to those 10 years of age and older. This restriction limits the utility of an age-based stratification in that some of the more interesting strata are not part of the sample.

Socioeconomic Status (SES)

Socioeconomic status (SES) offers another intriguing potential stratification variable. Indeed, within the Environmental Protection Agency, the concept of Environmental Equity and Environmental Justice is receiving a substantial amount of attention, given the perception that individuals from lower socioeconomic classes as measured by income or education and minority communities receive an unfair burden of pollutant exposures. This concept would suggest that a measure of SES, for example, family income, might be an appropriate stratification variable for NHEXAS investigations.

The hypothesis of inequitable distribution of environmental pollution, although well-established for lead poisoning, is untested for many pollutants. For exposure to other heavy metals, agricultural pesticides, and other NHEXAS classes of chemicals, the hypothesis has not been tested and certainly not proven. Given the embryonic nature of the science of exposure assessment, stratification based

upon SES-related variables is not yet warranted. Gathering of such information under the NHEXAS umbrella is essential and necessary; stratification based upon such hypotheses is not.

Although SES factors are not stratification variables, Environmental Justice and Environmental Equity issues will be addressed to some extent through stratification of political and census units by land use (urban, suburban, and rural) and by racial characteristics (preponderantly white and preponderantly minority) as described below in the section on geographic stratification.

Geographic Stratification

Geographic stratification is a third potential method of reducing variance in the sample means. Arguments related to this potential stratification are similar to those associated with SES stratification with the exception that measured pollutant values in various geographic areas (or strata) are different. Urban areas are noted for urban pollution: vehicular exhaust and concomitant secondary pollutants, industrial pollutants, hazardous waste sites, etc. Rural areas are noted for differing mixes of pollutants: relatively pristine air but often contaminated water due to agricultural runoff. Further, sources of food differ. In urban areas, most food is shipped in with only garden-grown produce being locally-produced. In rural settings, much of the produce may be locally-grown.

Geographical stratification offers an additional attractive feature in that some component of SES stratification can be included. Census information can be used to assess general income levels and racial make-up of populations down to the census tract level. We can make use of these data to ensure that a stratification based upon geographical region will also include, for example, members of racial or ethnic minorities or groups of varying SES.

Stratification in the NHEXAS Study Design

In this investigation, we will exercise geographic stratification. Three strata have been selected: urban, suburban, and rural. We expect that such a stratification is appropriate under Reason 1 above and that Reason 3, interest in these specific strata, also adds to the strength of the argument. Further, we expect to gather data on SES-related variables and on participant age that will afford some of the information accorded to other potential stratification variables.

1.2.4 Number of Cycles

An important consideration in a repeated-measures design is the number of repeated sampling Cycles. Two considerations come into play: statistical issues associated with improved statistical power (see Appendix 3, Statistical Issues Related to NHEXAS Design), and external, physical driving forces that affect variability.

Many pollutant exposures are driven by meteorological cycles. Air pollution transport is certainly affected by mesoscale meteorology. Snow cover, prevailing wind directions, rainfall, and numerous other meteorological parameters affect potential exposure for many media. Seasonal changes affect agricultural and industrial sources as well.

The length of such cycles varies. Synoptic meteorological cycles have a length of approximately two weeks. This is the length of time between successive cycles of low pressure followed by high pressure weather systems moving across the continent. Seasonal changes are linked with astronomical cycles, e.g., solstices and equinoxes, resulting in large-scale temporal variability in weather (e.g., summer versus winter). Other natural cycles could be used (e.g., long-term weather patterns governed by sunspots and *El NiZo*), but most are of duration longer than one year and cannot be taken into account in an investigation of 12-15 months duration.

The seasonal time-scales offer the most natural interval for exposure measurement. Activity patterns are influenced by local meteorology in a manner analogous to the changes seen in pollutant

concentrations. A measurement protocol designed to take one measurement each season would, in some sense, cover all possible contingencies. A modification of this protocol would allow two measurements during each season, since there are significant changes in meteorology and activity patterns on time scales of roughly six weeks. Such a protocol would provide a sense of the "within-season" variability of individual exposure.

We will implement the study using eight sampling Cycles. This is a compromise solution, but one with good arguments in its favor. Appendix 3, Statistical Issues Related to NHEXAS Design, outlines further arguments in favor of this sample design.

1.2.5 Sample Size Considerations

The optimum number of individuals to be monitored in a given Cycle is influenced by two factors: statistical and logistical.

Statistical Issues

The primary statistical issue associated with any sample to be drawn addresses the question of utility of the data with respect to drawing conclusions. A sample size that is insufficient will not yield precise estimates of, say, mean exposures experienced by the population. We are charged with establishing a sample size sufficient to ensure precise estimates of population parameters, but not so large as to result in logistical problems. (Sample design issues are discussed in detail in Appendix 3, Statistical Issues Related to NHEXAS Design.)

In many small exposure assessment or epidemiological studies, a sample size of approximately 50 has been deemed sufficient. In these investigations, a tacit assumption has been made: the data are normally-distributed (or can be transformed to a normal form). For a random sample drawn from such a distribution, a stable relationship between the mean and standard deviation is established. Thus the standard error for the mean estimate decreases as the square root of the sample size increases. Smaller sample sizes can have similar properties if the distributions are not wildly skewed. For example, some studies have used a sample size of 30 as sufficient to fulfill this criterion.

The final statistical issue that must be addressed is the potential misclassification induced by stratification. Implicit in the stratification itself is the assumption that such a variable will affect the outcome, in this case exposure to certain pollutants. Efficiency is lost relative to a totally random sample when an improper stratification is placed upon a sample. A good design must account for this contingency. The total sample size must be sufficient to ensure adequate description of population parameters even if such efficiency is lost by improper stratification. (Sample size issues are discussed in detail in Appendix 3, Statistical Issues Related to NHEXAS Design.)

Logistical Issues

Logistical issues are also important in the design of an investigation. Management of equipment, personnel, data, and costs must come into consideration. The overriding issue must be the quality of the results to be obtained. Total cost is a secondary issue. On the other hand, if a given design produces quality data but at a cost that cannot be made acceptable, one is at an impasse.

1.2.6 Final Survey Design

A reasonable and prudent study design calls for taking all of the above discussion into account. First we will assess the statistical issues associated with the survey design. We include the expected variability in samples measured, address issues associated with stratification, and study the most appropriate number of samples to be taken during the year at a given location. From this we develop a

sample size. Finally, we establish a cost for this sample size to be completed and assess the feasibility of carrying out the investigation given the funding available.

The collaborating organizations will collect environmental samples and questionnaire data to determine environmental concentrations in and around the residence, questionnaire data to characterize activities that bring respondents into contact with target chemicals, and biological samples to confirm that an uptake of target pollutants has occurred and to provide a relative measure of total exposure at home and away. Data will be collected using a series of questionnaires, a detailed in-home pollutant measurements protocol, and in-home food collection. Samples and questionnaire data will be collected by field staff, and analyzed and stored by the collaborating organizations. The information collection will involve at least 50 households, each including a target individual, to be studied during eight seven-day Cycles over the course of a year. Each Cycle provides 24-hour integrated samples of indoor and outdoor air taken over a 7-day period. Other media are sampled with grab or spot samples. The list of samples to be gathered is in Table 3 in Section 2.1.2.

The results will provide information on whether a single visit to a residence will supply adequate information to estimate annual average exposure, and what media and types of samples must be collected to ensure this prediction's accuracy. EPA Office of Research and Development (ORD) and the collaborating organizations will use the information to improve the design of the full-scale NHEXAS survey.

Hypothesis 1: Duration

- Pollutant exposures estimated from short-duration measurements are equivalent to actual exposures experienced over much longer durations (e.g., one year).

To test hypothesis 1, data collected in each Cycle will form a sub-dataset. Statistical tests will be used to compare the eight sub-datasets. If the data are normally-distributed, or can be transformed into normally-distributed (e.g., lognormally-distributed) standard, one-way analysis of variance techniques will be invoked to ascertain whether the Cycles differ from one another. Then the data for each individual will be analyzed similarly. If temporal variability is larger than or comparable to population variability, multiple-period monitoring will be necessary for the full-scale NHEXAS project. Two-way analysis of variance (ANOVA) will be the prime statistical procedure testing the relative variance between individuals and season. If the data are not normally distributed, non-parametric tests will be used for example, Kruskal-Wallis analysis.

Analysis will proceed on three levels. The first level of analysis is univariate. Each of the parameters measured (indoor air, outdoor air, etc.) will first be treated separately. All analyses will be done for each pollutant monitored. The results of this investigation will indicate the variability for each of these measurements over the 12-15 month sampling period. The second level of analysis will make use of the biological data as the dependent variable in a generalized linear model. Independent variables will include all the media concentrations as well as questionnaire information. From this approach, we can determine which media are most closely related to the biological measures of exposure. Further, through both within-Cycle (for the population) analysis and between-Cycle (for individuals) analysis we can assess the varying influence of such media over the 12-15 month period.

The third level of analysis involves estimating the total multimedia exposure for each pollutant. This will be effected using modeling techniques (MacIntosh et al. 1994a, MacIntosh et al. 1994b) designed to take such environmental data, questionnaire data, and activity profiles and produce and produce estimates of total exposure and total absorbed dose. Total exposure will be expressed in units of : g/day and will be calculated as the sum of the product of exposure concentration and contact rate (e.g., tap water concentration multiplied by water ingestion rate) for each exposure medium. Total

absorbed dose will be expressed in units of : g/kg/day and will be computed as described for total exposure, although each medium-specific exposure will be multiplied by the appropriate absorption coefficient or rate. Analysis of this newly-synthesized will be completed using similar techniques to those discussed in the previous paragraphs.

Hypothesis 2: Prediction

- Questionnaire data are sufficient to afford prediction of exposures experienced by a study population.

To test hypothesis 2, we will first construct appropriate exposure variables. As discussed earlier, two techniques can be envisioned. The first makes use of every potential distribution (117 in all). The second constructs a single total exposure estimate by combining information from time-activity patterns and some questionnaire data to obtain a single value. Analysis will proceed on both data sets. Questionnaire data will be used to categorize individuals according to characteristics of exposure. These may be demographic or housing characteristics, or they may use categorizations based upon activities (hobbies, work environments, etc.). If such categorizations are successful in placing respondents in differing exposure groups, monitoring burden on respondents can be reduced by noting that exposures are higher (or lower) given certain characteristics obtainable by questionnaire. If questionnaire data are not able to place respondents in differing exposure groups, then this component of the investigation may be reduced to simple demographic response data needed to ensure adequate representation of the population. The biological data will be treated as the dependent variable in a generalized linear model. Independent variables will include media concentrations and questionnaire data. Ability to predict biological concentrations from surrogate measures will be tested.

Hypothesis 3: Food

- Questionnaire data are sufficient to afford prediction of exposures experienced by a study population.

To test hypothesis 3, the 400 samples of duplicate-plate food will be analyzed. Mini-market basket food will also be analyzed. If the analyses are comparable, the diary method may be chosen because it is less burdensome to the respondents.

Reducing Burden and Cost

Additional data are to be collected via questionnaires in order to evaluate methods of reducing burden and cost in Phase II. For example, we will identify questions in the questionnaire that are useful in predicting urine and blood concentrations. We will also evaluate the ability of a combination questionnaire and mini market-basket approach in determining the component of total exposure coming from food. In doing so, we may reduce the burden to the respondent in maintaining a duplicate plate for diet contamination studies.

1.2.7 Pretest Studies

Prior to commencing the full-scale investigation, several short-duration pretest studies have been undertaken in the homes of the Principal Investigator and Harvard University staff. The purpose of these investigations is to work out protocol details, not to gather data to be analyzed for temporal

trends. For example, we have worked out the duties of the field staff necessary for the study:

Field Interviewer:	administration of questionnaires and diaries; collection of water, food, dermal wipes, and one urine sample
Field Technician 1:	air sampling equipment set-up and take-down
Field Technician 2:	soil and dust sampling, house and yard plans, technician questionnaire
Phlebotomist:	collection of blood, personal monitor, and one urine sample

For more detail on field staff duties, see Section 1.3, Personnel Qualifications. In our pretest studies we tested this division of labor, measured the amount of time necessary for each operation, and assigned tasks accordingly. For more detail, see Section 2.1.5, Sample Collection Time Estimates.

Another purpose for the pretest studies is finalization of the Technician's Reference Manual. This manual will contain all Standard Operating Protocols for the field, including the routine to be followed while in the home. A list of the SOPs is in Section 3.5, Procedures.

Pretesting the NHEXAS Questionnaires is being carried out by another group. We will make use of their results in our final field protocol.

1.3 Personnel Qualifications

The Principal Investigator and other coordinators and scientists have appropriate degrees and experience; see Section 2.3, Project Responsibilities.

Qualifications for laboratory technicians include:
- a bachelor's degree in a scientific, technical, or medical/social field; chemistry is preferred
- laboratory experience; analytical experience is preferred

The field team for the NHEXAS Phase I Field Study (Field Interviewer, Field Technicians 1 and 2, and Phlebotomist) will be trained professionals from Westat. Qualifications for field staff include:
- a bachelor's degree in a scientific, technical, or medical/social field; chemistry or environmental engineering is preferred for technicians
- experience in field research is preferred
- some field staff should have a second language appropriate to the population being surveyed
- the phlebotomist will be licensed by the appropriate state agencies

The Interviewer will be responsible for:
- administration of descriptive, baseline, and follow-up questionnaires
- instructing the target individual how to complete the time-activity diary and the food checklist, and checking the finished diaries for completeness
- instructing the target individual on the use of the personal air sampler
- preparation and maintenance of equipment used for collection of dermal wipe, water, duplicate diet, and urine samples
- collection of dermal wipe, water, duplicate diet, and Day 8 urine samples

Field Technician 1 will be responsible for:
- setup, calibration, maintenance, and takedown of air sampling equipment
- collection of indoor and outdoor air samples, and setup of personal air sampler

Field Technician 2 will be responsible for:

- drawing plans of the house and yard
- sampling dust and soil
- completing the Technician Questionnaire

The Phlebotomist will be responsible for:
- all blood sampling procedures including apparatus preparation, storage of samples, and record keeping associated with such sampling
- retrieval of personal air sampler
- collection of Day 2 urine sample

The Field Coordination Center (FCC) staff consists of the Field Coordinator, the Field Supervisor, the Communication Specialist, three Resource Specialists, the FCC Technician and assistant, the FCC Clerks, and the Telephone Interviewers.

1.4 Training Required

Harvard and Emory staff will train the laboratory technicians who will perform analyses of air particulate, dust, soil, and dermal wipe samples at Emory's Trace Metals Laboratory. They will learn the specific analytical techniques to be used, particularly use of the GF-AAS instrument. They will also be trained in the use of the ICP-MS instrument, which may be employed for confirmatory analysis. The training is described in more detail in SOP G09 "Training of Laboratory Technicians."

Responsibility for training the field staff will be joint between Emory, Harvard, and Westat. Westat will take the lead in training the interviewers. This training will take place at Westat's corporate offices in Rockville, Maryland. During a six-day training period, the interviewers will learn:
- standards of behavior, conversation, and appearance
- all aspects of the questionnaires and diaries, and how to administer them effectively
- techniques for collection and handling of dermal wipe, duplicate diet, and urine samples
- how to mark any data or samples that may have problems

The training is described in more detail in SOP G08 "Training of Interviewers."

Training for the field technicians will take place at Westat. Training will be by Emory, Harvard, and Westat engineers familiar with all aspects of the field instrumentation. During a six-day training period, the field technicians will learn:
- standards of behavior, conversation, and appearance
- sample handling techniques including labeling, logsheets, forms, and storage requirements
- preparation of equipment and containers for field visits
- setup, operation, and disassembly of field sampling equipment
- choosing appropriate sampling sites
- all aspects of the field technician questionnaire, and how to administer it effectively
- how to collect and handle air particulate (indoor, outdoor, and personal), dust, soil, and water samples
- how to mark and handle any data or samples that may have problems

The training is described in more detail in SOP G07 "Training of Field Technicians."
Training for phlebotomists will take place at Westat. Phlebotomists will learn:
- standards of behavior, conversation, and appearance
- sample handling techniques including labeling, logsheets, forms, and storage requirements
- preparation of equipment and containers for field visits
- procedures for retrieval of the personal air sampler, including measuring flow readings

The training is described in more detail in SOP G10 "Training of Phlebotomists."

Training for field interviewers, technicians, and phlebotomists will be supplemented by actual field sampling in residences occupied by Westat staff members. Trainees will be tested during training and at the end of training. During the early part of the survey, instructors will accompany field staff to evaluate their performance and provide any help needed. During the 15-month survey period, internal audits every six months and a performance audit by an independent entity will be performed.

The Field Coordination Center staff will receive training appropriate to their duties. The training is described in more detail in SOP G11 "Training of Field Coordination Center Staff."

1.5 Experimental Design

Data to be collected in this investigation include:

- questionnaire data describing factors likely to influence environmental exposures, activity pattern data, environmental concentrations, and biological data
- environmental data from analysis of samples of air particulates, soil, dust, water, and food
- biological data, assumed to represent the actual exposures experienced by the respondents

The questionnaire and environmental data will be assessed as to their ability to predict the exposure variables. Thus the biological variables will be considered dependent variables and the questionnaire and environmental variables the independent variables.

Several types of analyses are envisioned. In particular, statistical trends in the means for the various environmental variables will be examined. Distributions will be determined at each sampling Cycle with the mean, median, standard deviation, and other statistics being calculated. Analysis of variance (ANOVA) will be used at each Cycle to evaluate the differences in urban, suburban, and rural strata. A comparison will be made across Cycles to determine whether the temporal variability is larger than the variability in the population for a given Cycle. This will first be attempted using simple t-tests on the means with comparisons made to the standard deviation in the mean exposure estimate. This will be followed by a two-way ANOVA using Cycle and individual as the variables. If the inter-individual variability is large in comparison with the temporal variability, this suggests that a single monitoring period is adequate to determine long-term population variability. If the converse is true, this would suggest the need for multiple samples over the course of one year.

Acceptable limits on false positives and false negatives will be:

$$\alpha = 0.05 \qquad \beta = 0.20$$

In other words, no more than five percent of positives may be false, and at least eighty percent of negatives will be true. The significance level will be 95%, and the power will be 80%.

The primary use of the data collected in this study will be to test the first hypothesis, regarding prediction of annual average exposures from short-term concentration data. Data will be gathered on surrogate measures for exposures, including in-home air, water, dust, food, and soil pollutant concentrations, as well as direct measures of exposure indicated by biological measurements.

To address the first hypothesis, we will use statistical methods to assess the ability of surrogate measures to estimate the biological concentrations: first for short-term estimates, then for long-term estimates gathered over the year-long study. We will test the accuracy with which measurements of metals, pesticides, and PAHs in environmental and biological media can predict annual average concentrations and, ultimately, annual average exposures.

Questionnaire data collected concurrently will be used to estimate exposures. If questionnaire data are not sufficient, we will couple that information with the in-home pollutant concentration to evaluate the improvement such data collection would give. In this manner, we can determine the minimum number of home visits and samples -- thus the minimum burden and cost -- needed for Phase II to give good exposure estimates.

2.0 PROJECT IMPLEMENTATION PLAN

2.1 Project Design Criteria

2.1.1 Criteria for Siting, Sampling Interval, and Number and Selection of Respondents

Overview

Siting Criteria: Sampling for the NHEXAS pilot investigations will take place in Maryland. Geographic stratification into urban, suburban, and rural strata will be effected. The urban stratum will be in the city of Baltimore. Suburban and rural strata will be developed using Westat's capabilities in survey design (see below).

Sampling Interval: The Harvard/Emory group protocol has been developed with an eye toward estimating annual average exposures from short-term (e.g., one-week or shorter) pollutant concentration measurements. To simulate results obtained from the other NHEXAS investigations, short-duration sampling periods have been selected to match those of the RTI/EOHSI consortium and the UA/IIT/Battelle consortium. These will be repeated over a 12-15 month period. We propose eight repetitions or Cycles with approximately six weeks between the beginnings of consecutive Cycles.

The six-week interval was chosen as a compromise. (See Appendix 3, Statistical Issues Related to NHEXAS Design.) Ideally, if the sampling period is seven days, one would prefer to sample consecutive seven-day periods throughout the year, a total of 52 periods. From a cost perspective, this is not feasible. Additionally, data collected in previous investigations (Quackenboss, et al., 1986, Spengler, et al., 1994) have indicated strong autocorrelation in pollutant measurements; sampling consecutive weeks would collect data of questionable value. Alternatively, sampling substantially less frequently, say two or three times over the course of one year, would leave large time gaps unmonitored. A protocol calling for one monitoring period each season was investigated but also deemed too sparse with respect to potential data use. Inspection of meteorological records for the U.S. suggests that real climatic differences have a spacing of about six weeks; seasons and weather conditions are substantially different with a spacing of about this increment. As local meteorological differences have significant impact on exposures, the Harvard group has adopted this interval between sampling Cycles. Additionally, monitoring twice per season will afford a better understanding of the variability in exposure patterns within seasons.

Number of Respondents: Upon selection of the six-week monitoring Cycle, the actual number of respondents monitored during each Cycle was determined, in part, by the funding level provided. Additional considerations included the need for each potential stratum to contain a sufficient number of respondents to allow statistical comparisons to be made between strata. Also, the total number of respondents must be sufficient to afford comparisons between Cycles for the entire distribution. Given the conflicting constraints of funding level and the need for data, a final figure of 50 respondents in each Cycle was determined. These 50 respondents are to be divided among three strata: urban, suburban, and rural. In order to attain 50 completed cases (see below), more than 50 households will be recruited. For more detail, see below and Appendix 3, "Statistical Issues Related to NHEXAS Design."

Respondent confidentiality will be assured. See section 4.5 and SOP G02 "Procedures for

Ensuring Confidentiality of Respondents' Records." Respondents will be given sufficient information to consider whether to participate, and will sign an informed consent form approved by the HSPH Human Subjects Committee.

Detailed Sample Selection Procedure

Westat will select and recruit a sample of households according to the procedures described in this section. The target sample size is 50 completed cases. A case will be considered complete if the respondent completes at least 6 of the 8 Cycles. The initial sample size is determined by the target number of completed cases and estimates of the response rates at each stage of sampling.

Respondents will be selected and recruited by selecting a sample of households following the survey sampling procedures described below. This will provide valid data on the probable response rates for the full-scale NHEXAS survey.

Approaches to Maximize Completion Rates

There are several stages of sampling, from the initial identification and selection of households, to the completion of eight Cycles of data collection visits by the field teams. Each of these stages has opportunities for nonresponse. The causes of nonresponse include refusals by the respondent, inability of the respondent to provide the required data and samples, inability of the project staff to contact the respondent and schedule the visits, etc. It is possible for respondents to participate in some, but not all of the eight Cycles of data collection visits. In fact, we expect that a fairly large number of sampled households will not participate in all Cycles.

Environmental monitoring studies place large burdens on the respondents and, consequently, have achieved noticeably lower response rates than less intrusive surveys. Studies with low response rates have the potential for significant nonresponse biases. In addition, large respondent burdens sometimes lead to a diminishment of the quality of the data provided by the respondents.

A number of methods have been used to induce respondents to cooperate. They generally fall into one or more of the following categories of inducements:

1. Appeals to authority, patriotism, virtue (e.g., help save the environment), altruism (e.g., help the common good or help advance science);
2. Reports about the pollutants found in respondent's home (this may be a counter-incentive to people who don't want to know; it may also raise liability issues, especially in rented homes);
3. General information about the pollutants -- sources, pathways, and effects; how to reduce exposure; how to obtain more information; where to go for help; and
4. Tangible incentives such as money (fixed amounts, variable amounts, lottery tickets), trinkets (tee shirts, caps, buttons, mugs, etc.), or "useful" gifts, perhaps related to the study such as environmentally safe substitutes for common items containing the pollutants (non-polluting paint, safe adhesives, praying mantis eggs, etc.)

A number of researchers have designed and conducted experiments to test the utility of incentives, and to estimate the increase in response rates associated with incentives. A few are summarized here. In the first Health and Nutrition Examination Survey (HANES I), conducted in the early 1970s, a $10 incentive was associated with a statistically significant 12% increase (from 70% to 82%) in the response rate over no incentive. All subsequent HANES studies have incentive payments. In HANES II, a $20 incentive was found to significantly increase the response rate (from 74% to 79%) over the $10 incentive used in HANES I. In the national adult literacy survey, respondents were interviewed and given a test to measure their reading ability. An incentive of $20 significantly improved the response rate over no incentive, especially among Blacks and Hispanics. However, a $35 incentive

was not significantly better than the $20 incentive. Further, the use of an incentive actually reduced the cost per respondent, because fewer callbacks were required to obtain a completed case.

The large respondent burden in this study requires a commensurate respondent incentive. Accordingly, we propose that an incentive be given at each Cycle of data collection visits. We propose a large incentive, $100, be given on each of the first four Cycles that a respondent completes, to establish and maintain good will with the respondent, and to compensate the respondent for the costs incurred in collecting the duplicate diet. To give the respondent an incentive to remain steadfast through the last four Cycles, we propose $150 for each of these Cycles. A respondent who completes all eight Cycles would therefore receive $1,000. If a respondent provides only one biological medium (blood sample or two urine samples) the payment will be reduced by $25. Any respondent who does not provide either a blood sample or two urine samples will not be paid for that Cycle. Respondents will also be reimbursed $60 per Cycle for food. While these respondent incentives are large, we believe they are necessary to induce respondents to consent to participate in the study and to remain with the study throughout it.

Estimation of Completion Rates for Survey Design

For the purposes of this study, we will define a completed case as a household that participates in six or more of the eight Cycles. This definition control the level of effort required to achieve an acceptable number of cases with an acceptable number of Cycles of data.

To estimate response rates, one typically looks for similar prior studies for their reports of response rates. There are no prior studies exactly like this one. However, there are some recent studies that have a number of features in common with this NHEXAS pilot. The overall response rates for the NOPES study and EPA TEAM-VOC study were 40% and 44% respectively (among those eligible after screening). These studies imposed burdens similar to the proposed study, but they had only one cycle of data collection visits, not eight. NHANES III asks respondents to undergo thorough medical and dental examinations at mobile examination centers. The examinations include biological samples. To date, nearly 30,000 persons have been sampled; nearly 80% of all respondents and over 80% of the Blacks and Hispanics have been examined. A recent study of nitrogen dioxide exposure in domestic microenvironments in Boston (Ryan, et al., 1988a, Ryan, et al., 1988b) investigated seasonality through three visits to respondents' homes over the course of a year. Of the 973 sampled and eligible households, 60% agreed to participate. Of these, 89% were monitored at least once and 76% were monitored all three times. The attrition rate between visits was between 6% and 13%. The Boston NO_2 study had three visitations per household, but the respondent burden per visit was measurably smaller. This respondent burden differential would suggest that the present study should have somewhat lower response rates than the Boston NO_2 study. We will use a full array of techniques, including substantial respondent incentives, to help increase the response rates. Accordingly we will assume, for design purposes, that t study will achieve response rates slightly lower than the Boston NO_2 study.

Specifically, we have made the following response rate assumptions. (See Appendix 3, Statistical Issues Related to NHEXAS Design, for sensitivity analysis.)

In the first Cycle:

- 25% of the households selected e initial sample will be successfully contacted and will consent to participate in the study.
- 87% of the households who consent to participate will complete the first Cycle.

In the second through eighth Cycles:

- 87% of the households who participated in the previous Cycle will participate in the current Cycle.
- 16% of the households who did not participate in the previous Cycle, but did participate in the

one before that, will participate in the current Cycle
- 2% of the households who did not participate in either of the two previous Cycles, but did participate in the one before that, will participate in the current Cycle.
- None of the households who did not participate in any of the three previous Cycles will participate in the current Cycle.

These assumptions imply that 50.6% of households that consent to participate will result in completed cases (six or more completed Cycles). If n is the number of households that consent to participate, then the expected number of completed cases is $.506n$. Setting this equal to the target number, 50 completes, yields the required number, 100, of households consenting to participate, which in turn, requires 400 households in the initial selection.

The model described above also implies that 488 data collection visits will be completed, 88 more than would be required if there were no attrition.

It is important to note that these response rates and sample sizes are projections, only loosely grounded in applicable data. The projected number of completed cases is an estimate, not a guarantee. Westat will follow current best survey research practices, including the procedures and protocols described here, and we will make our best efforts to achieve the highest possible completion rates. If, after Cycle 1, the number of households agreeing to participate in Cycle 2 is lower than expected, more households could be recruited. If, despite our best efforts, the actual response rates turn out to be lower than the assumed values, then fewer than 50 completes may be realized.

Household Selection Procedures

The sample will be constructed to achieve:
- Adequate representation of urban, suburban, and rural households
- Adequate representation of both white and minority households
- Highest possible completion rate
- Good survey research practices throughout
- Testing of sample selection and survey research procedures that could be employed in the full-scale NHEXAS study
- Minimum field costs

Westat maintains a master sample of 100 geographically-defined county clusters representing the 50 states and the District of Columbia. The master sample was designed using 1990 Census data on population, race, ethnicity and 1988 Bureau of Economic Analysis data on per capita income. A statistically valid area probability sample of homes in central Maryland has been designed and selected. As discussed elsewhere, representation across the urban-suburban-rural gradient is desired. To achieve this, we have selected the Baltimore MSA, to provide the urban, suburban and some of the rural areas, and a contiguous county, Talbot, that is within Maryland, but not included in any MSA. A county contiguous to the Baltimore MSA was selected in order to control the costs and time required to travel there. Talbot County will provide additional rural areas for the sample.

Within these counties, Census block groups were stratified into

1. Urban, predominantly white
2. Urban, predominantly minority
3. Suburban, predominantly, white
4. Suburban, predominantly minority
5. Rural

This was accomplished using the Census file STF-3A, "Census of Population and Housing, 1990: Summary Tape File 3A on CD-ROM", which provides data on a number of characteristics, including

urbanization and race for Census blocks nationwide. A Census block group is a cluster of Census blocks within a Census tract. Census tracts and, consequently, block groups never cross county lines. Block groups generally contain between 250 and 550 housing units.

A Census block group was defined to be predominantly white if the percentage of White residents exceeds 50 percent. Otherwise it was considered predominantly minority.

A Census block group has been defined to be urban if it is in Baltimore City, or is in Baltimore County or Anne Arundel County and has a population density of 10,000 or more persons per square mile. It was considered rural if it is in Talbot or Queen Anne"s Counties, or is in Baltimore County or Anne Arundel County and has less that 25% urbanized areas and less than 1,000 persons per square mile. All remaining block groups in Baltimore County or Anne Arundel County were considered suburban. An urbanized area is defined by Census as comprising one or more places and the adjacent densely settled surrounding territory "urban fringe" that together have a population of at least 50,000 persons. The urban fringe generally consists of contiguous territory having a density of at least 1,000 persons per square mile. It is to be noted that under these Census definitions, an MSA can have rural areas within its boundaries and non-MSA places can be suburban or even urban.

After the block groups within the target counties were stratified, five block groups were selected from each stratum with selection probabilities proportional to their size (PPS), with the size measure being the number of residents.

At this point in the development of an area probability sample, the usual procedure is to list all households in the sampled block groups. In order to save project resources for the actual data collection, Westat employed an alternative approach to the next stage of selection. Lists of residential telephone numbers were obtained from a commercial vendor and households were sampled from these lists. Simple random, equiprobable sampling was used at this stage. Commercially available telephone lists are generally based on telephone exchanges. Since the boundaries of telephone exchanges and block groups do not coincide, this procedure introduces some noise into the study. However, we judge that this error will be negligible, relative to other error sources in the study.

For the initial selection, 20 households will be selected from each tract, for a total of 500 households. This is larger than the 400 we expect to need, to provide extra households available if needed. The target number of completed cases is two households per tract. This target leads to the following distribution of households in the sample:

20 urban households	30 white households
20 suburban households	20 minority households
10 rural households	

Recruitment Strategies

The targeted completion rate for the NHEXAS Pilot study is to complete 50 households over a series of eight Cycles. In order to maximize the recruitment and retention for the 50 completed households, the HSPH/Westat consortium proposes the following recruitment and retention strategies based on the two different incentive types:

The consortium will generate as much advance publicity on the study as possible within each community. Letters will be sent to the leaders of each community such as the mayor of the town, county health department, etc. Press releases will be prepared and mailed out to the local newspapers in each community. The public outreach team will also obtain as many endorsement letters as possible from public officials, community leaders, and representatives of civic and community organizations. Some of the endorsers will be photographed wearing the monitoring equipment and otherwise participating in the study. Endorsements will be sought from:

- U.S. Sen. Barbara Milkulski

- Baltimore Mayor Kurt Schmoke
- County Executives in the target counties
- Mayors of the larger cities and towns in the target counties
- Officials from local health departments
- JHU Dean of the School of Public Health

The HSPH/Westat consortium will send an introductory letter to the sampled households to introduce the study, and begin the process of recruiting respondents. Themes in the letter include: this is an important environmental study; you are fortunate enough to have been chosen; you represent thousands of your neighbors and fellow Americans; it is important for you to participate; and you will be telephoned soon.

After the letters have been mailed, telephone interviews will be conducted in order to recruit participants. The descriptive questionnaire will be administered in order to enumerate the household, and a household member will be selected to participate in the study. We will use the same method as RTI, to maintain consistency of methodologies across consortia. The interviewer will set up an appointment for the field team to visit the household to administer the baseline questionnaire and to collect the appropriate samples. If a potential participant is unsure about what is involved in the study and whether to participate, the telephone interviewer can set up a household visit for a representative of the field team to make a demonstration of the type of equipment that will be used. The interviewer will have copies of the endorsement letters and other promotional and descriptive material to show to the respondent, and will give the propsective respondent a token gift as a respondent incentive: options under consideration include mugs, movie passes, restaurant discount coupons.

The recruitment will offer the following incentives for participation:

-- Provide results on findings which include information on the possible contaminants in the home, and possible health effects to the individuals and their children;
-- Offer the results of the water analysis and tests for lead levels and pesticides;
-- Award monetary incentives which are described below;
-- Assist in the benefit to the advancement of science and in measuring the contaminants to the environment.

After any recruitment contact, each participant will receive a thank you postcard.

Retention Strategies

Security and confidentiality issues will be addressed with each household. A fact sheet with answers to commonly asked questions will be prepared and will be provided to each field team.

Extensive interviewer training will be provided for the approach on entering the household and the appropriate behavior and presentation once in the household. Extensive refusal conversion training will also be provided to every member of the field team to ensure that all of the potential concerns of the household member will be addressed.

Promotional items will be provided to the selected household member at various phases throughout the study. Some of the gifts will correspond to samples being taken. One gift will be given per cycle, without previous announcement. Most of the gifts will contain the project logo, the 800 help line phone number, and/or reminders of activities or samples the respondent is to provide, e.g., the duplicate diet, urine samples, etc. Previous research has indicated that gifts such as these help to persuade the respondent to respond. In addition, with the help line phone number, and reminders, concern for the respondents concerns is shown and we are making it easier for him/her to participate. Specific gifts currently under consideration include:

-- Blood sample - small box of cookies and/or juice
-- Activity diary - pencils, note pads
-- Soil/air samples - bulbs or plants; button

-- Water sample - tumbler or sports cup
-- Dermal wipe - hand lotion samples; small towel to dry hands
-- Duplicate diet - plastic measuring cups or a scoop with the study logo; insulated lunch bag; rubberized jar opener; refrigerator magnet; bag clips; restaurant discount coupons
-- Dust sample - dust absorbent cloth with

Individual Selection

After developing a roster of those living within the home, a random individual will be selected as the target participant. This procedure will invoke a random number generator to select the individual from the roster. Every effort will be made to solicit participation by the target individual. If the target individual refuses to participate but another individual within the residence is willing to carry out the tasks associated with the study, that second individual will be enrolled. It is felt that, because most of the measurements are household based, this method offers the best alternative and ensures that a "willing" household is not lost to the investigation due to an "unwilling" individual participant.

2 Sample Collection

Environmental samples will be collected and analyzed as indicators of the environmental exposures of an individual. The pollutant groups under consideration in this investigation are toxic metals, pesticides, and polycyclic aromatic hydrocarbons (PAHs).

The collection of samples will be carried out following standard operating procedures (SOPs). (Appendix -- Standard Operating Procedures for the Harvard/Johns Hopkins NHEXAS Investigations). Table 3 outlines the sampling protocol for the proposed investigation. Section 2.1.3 gives more detail on protocols. The SOPs (listed in Section 3.5) give specific information concerning the sequencing and frequency of monitoring. Each sampling Cycle consists of three visits on days 1, 2, and 8.

Table 3 " Sampling Protocol for Harvard Group NHEXAS Pilot Studies

Medium	Description	Day(s) Taken	Number of Samples, Analytes	Notes
Indoor Air	24-hour sample over 7 days. 4 L/min pump with timer.	1-8	1: metals 1: pesticides/PAHs	Single pump with timer and switching apparatus for metal and pesticide/PAH samplers.
Outdoor Air	24-hour sample over 7 days. 4 L/min pump with timer.	1-8	1: metals	Pump with timer.
Personal Air	24-hour sample, 4 L/min. Target individual only.	1-2	1: metals	Small Personal Exposure Monitor (PEM).
Water	Drinking & cooking water sample, flushed (2 min) if tap water.	1	1: metals 1: pesticides/PAHs	More representative and less burdensome than first daily draw.
House Dust	Vacuum sample on measured floor areas of 3 rooms.	1	1: metals, pesticides/PAHs	Used with soil and dermal samples.
Soil	Composited sample from yard & play area, food garden, near foundation.	1	1: metals, pesticides/PAHs	Used with house dust and dermal sample to evaluate impact of soil on exposure.

Dermal Wipe	Target individual, both hands.	1	1: metals	Used with soil and house dust samples.
		8	1: pesticides/PAHs	
Duplicate Food	4-day home composite, target individual only.	3-6	2 (solids & beverages): metals, pesticides/PAHs	Used with mini-market basket approach.
Urine	First daily void, target individual only.	2	1: metals	First daily void less burdensome than 24-hour sample.
		8	1: pesticides/PAHs	
Blood	56-cc venous puncture, target individual only.	2	7: metals, pesticides, PAHs	Taken by licensed phlebotomist.

Table 4 shows the field staff responsibilities during each visit of the sampling Cycle.

Table 4 -- Field Staff Responsibilities and Schedule

Visit and Day	Interviewer	Field Technician 1	Field Technician 2	Phlebotomist
Visit 1: Day 1	Questionnaire: Baseline Instruction: Activity diary Food checklist Duplicate diet Urine samples Personal air sampler Samples: Water Dermal wipe - metals Complete logsheets and chain-of-custody forms	Setup for sampling: Indoor air - metals Indoor air - PAHs & pesticides Outdoor air - metals Personal air sampler Complete logsheets and chain-of-custody forms	Plans: House Yard Foundation Samples: Dust Soil Questionnaire: Technician Complete logsheets and chain-of-custody forms	
Visit 2: Day 2				Sampling: Blood Pickup: Personal air sampler Urine sample 1 Complete logsheets and chain-of-custody forms
Visit 3: Day 8	Questionnaire: Followup, Food Diary Followup Diary examination: Activity diary Food checklist Pickup: Duplicate diet Urine sample 2 Sample: Dermal wipe - pesticides Complete logsheets and chain-of-custody forms	Takedown: Indoor air - metals Indoor air - PAHs & pesticides Outdoor air - metals Complete logsheets and chain-of-custody forms		

2.1.3 Media Protocol Synopses

Indoor Air -- Metals: Indoor air sampling will be accomplished through the use of a modified Harvard Black Box Sampler. This apparatus operates at a flow rate of 4 L/min using a mass flow controller to ensure steady flow during sample collection. The apparatus is equipped with a sophisticated timer, controller, and solenoid valves that allow two inlets to be sampled on an intermittent time schedule. Flow will alternate through the metals sampler and the pesticide/PAH sampler.

A 24-hour sample will be taken over a 164-hour period (just under one week). The metals sampler

includes a single-stage impactor with a sharp cutpoint at 10 :m to provide a sample in the inhalable size range. The sample air volume target is 5.76 cubic meters. The sample collection medium will be cellulose ester membrane filters appropriate for trace metal analysis, taken from a single lot.

Filters will be extracted and analyzed in accordance with EPA Method 200.8 and/or 200.9. (See SOPs F02, L06, L07, and L08 for details.)

An additional 5% of the total number of planned exposure sample filters will be retained in the laboratory for use as lab blanks. An additional 10% will be used as field blanks; these will be carried through procedures identical to sample filters but will never be attached to an operating sample pump. All filters will be from the same lot.

Samples will be analyzed for concentration of arsenic, cadmium, chromium, and lead by atomic rption spectrophotometry.

Expected concentrations and limits of detection: see Table 5. The instrumental limit of detection for the given metal analytes is in the low parts per billion (ppb) range in solution.

Table 5 -- Expected Metal Concentrations and LODs for Air Sampling

Metal	Expected Concentration (ng / m^3)			Expected Limit of Detection Given Sampling Protocol
	Indoor	Outdoor	Personal	
Arsenic (As)	10	15	10	4.1 ng/m^3
Cadmium (Cd)	34	34	34	0.41 ng/m^3
Chromium (Cr)	38	22	38	0.8 ng/m^3
Lead (Pb)	33	26	35	6.0 ng/m^3

Outdoor Air -- Metals: Outdoor air sampling will be accomplished through the use of a sampling pump operating at 4 L/min. Using an integral timer-controller, a 24-hour sample will be taken over a 164-hour period. Outdoor sampling will include rain protection for the sampling head. Filter media, wet chemical workup, and analysis will be identical to the indoor sample.

Expected concentrations and limits of detection: see Table 5.

Outdoor air will not be sampled for pesticides and PAHs because studies such as the Nonoccupational Pesticide Exposure Study (NOPES) (Immerman and Schaum 1990) found outdoor pesticide levels to be very low, often below the limit of detection.

Personal Air -- Metals: A personal air sample will be collected from each target individual. This requires outfitting the target individual with a Personal Exposure Monitor (PEM) for one 24-hour period during the 7-day sampling period. The target individual will wear the sampler except when sleeping, bathing, swimming, or participating in contact sports; then it will be placed in a safe location nearby. The sampling apparatus will consist of a small personal pump operating at 4 L/min connected to a PEM with a 10 :m cutpoint; particulates are collected on cellulose ester membrane filters. Workup and analysis of such samples is similar to that described for the indoor air samples. (See SOPs F03, L06, L07, and L08.) All filters will be from the same lot.

Expected concentrations and limits of detection: see Table 5.

Indoor Air -- Pesticides and PAHs: Pesticides will be collected using a pump operating at 4 L/min operating over the 164-hour sampling period. The pumping system is as described above for Indoor Air -- Metals; the pesticide/PAH sampler is attached to the second air inlet. Approximately 5.76 cubic

meters of air will be sampled. Sample collection will be on quartz fiber filter and polyurethane foam (PUF). Extraction will be via Soxhlet extraction. Analysis will be by GC/MS. (See SOPs F02, L10, and L14.)

An additional 5% of the total number of planned exposure PUF sampler will be retained in the laboratory for use as lab blanks. An additional 10% will be used as field blanks; these will be carried through procedures identical to other samples but will never be attached to an operating sample pump. Blanks will be from the same lot as sample filters analyzed at the same time.

Expected concentrations and limits of detection: see Table 6.

Table 6 -- Expected Pesticide and PAH Concentrations and LODs for Indoor Air Sampling

Category	Pollutant	Expected Concentration	Expected Limit of Detection Given Sampling Protocol
Pesticides	chlordane	40-300 ng/m^3	12 ng/m^3
	chlorpyrifos	5-400 ng/m^3	26 ng/m^3
	4,4'-DDD	0 ng/m^3	1.8-5.3 ng/m^3
	4,4'-DDE	1-5 ng/m^3	1.4-3.6 ng/m^3
	4,4'-DDT	0.2-1 ng/m^3	2.2-4.1 ng/m^3
	dieldrin	1-10 ng/m^3	0.5-3.3 ng/m^3
	heptachlor	0.3-200 ng/m^3	0.5-3.1 ng/m^3
	malathion	1-20 ng/m^3	18 ng/m^3
PAHs	anthracene	3.0 ng/m^3	31 ng/m^3
	benzo(a)pyrene	1.0 ng/m^3	19 ng/m^3
	chrysene	2.3 ng/m^3	14 ng/m^3
	phenanthrene	18.0 ng/m^3	12 ng/m^3

House Dust -- Metals: House dust samples will be collected from measured floor areas in three rooms, using a vacuum cleaner with a standard cyclone. The sample will be collected in a Teflon jar and sent to a laboratory to be sieved and divided. Extraction and analysis of the metals fraction will be similar to those for the air particulate samples. (See SOPs F04, L05, L06, L07, and L08.)

Expected concentrations and limits of detection: see Table 7.

House Dust -- Pesticides and PAHs: Extraction and analysis of the pesticide/PAH fraction will be similar to those for the air particulate samples. (See SOPs F04, L05, L11, and L14.)

Expected concentrations and limits of detection: see Table 7.

Soil -- Metals: 24 cores will be taken outside the residence from the yard and play area, the food garden if any, and near the foundation of the house. Cores will be taken using plastic syringes cut and marked to a standard length. The composited sample will be sieved and divided. Extraction and analysis of the metals fraction will be similar to those for the air particulate samples. (See SOPs F05, L05, L06, L07, and L08.) Expected concentrations and limits of detection: see Table 7.

Soil -- Pesticides and PAHs: Extraction and analysis of the pesticide/PAH fraction will be similar to those for the air particulate samples. (See SOPs F05, L05, L11, and L14.)

Expected concentrations and limits of detection: see Table 7.

Table 7 --Expected Concentrations and LODs for Dust and Soil

Category	Pollutant	Dust			Soil	
		Expected Concentration (ng/g)	Expected Loading (ng/m^2)	Expected Limit of Detection* (ng/g)	Expected Concentration (ng/g)	Expected Limit of Detection* (ng/g)
Metals	Arsenic	500-25,000		350.	500-25,000	350.
	Cadmium	100-5,000		12.5	100-5,000	12.5
	Chromium	5,000-250,000		62.5	5,000-250,000	62.5
	Lead	10,000-500,000		75.	10,000-500,000	75.
Pesticides	Chlordane-"	400	530	12.14	600.	0.81
	Chlordane-(	460	550	14.30		0.95
	Chlorpyrifos	540	820	25.62	18	1.71
	4,4'-DDD			73.73	5	4.92
	4,4'-DDE	170	620	13.62	50	0.91
	4,4'-DDT	250	530	18.36	1-100	1.22
	Dieldrin	250	560	55.23	2-10	4.35
	Heptachlor	280	300	43.23	1-10	2.88
	Malathion	220	420	21	7	1.38
PAHs	Anthracene			100.	10-1,000	100.
	Benzo(a)pyrene	1120	1540	130.	2-1480	130.
	Chrysene	1260	1780	100.	2-1600	100.
	Phenanthrene			100.		100.

*Expected Limit of Detection Given Sampling Protocol

Dermal Wipe- Metals: The target individual's hands will be wiped with isopropanol-soaked wipes on day 1. Wet chemical workup and analysis for these samples will be similar to those for the air particulate samples. (See SOPs F06, L06, L07, and L08.) Dermal wipe samples will be used in conjunction with house dust and soil samples to evaluate the impact of local soil on exposure.

Expected mass per sample and limits of detection: see Table 8.

Dermal wipe -- Pesticides: Samples will be collected on day 8 using a protocol similar to that for metals. Extraction and analysis will be similar to those for the air particulate samples. (See SOPs F06,

L12, and L14.)

Expected mass per sample and limits of detection: see Table 8.

Table 8 -- Expected Concentrations and LODs for Dermal Wipe Sampling

Category	Pollutant	Expected Mass per Sample	Expected Limit of Detection Given Sampling Protocol
Metals	Arsenic	< 700 ng	700. ng
	Cadmium	< 25 ng	25. ng
	Chromium	< 125 ng	125. ng
	Lead	< 150 ng	150. ng
Pesticides	Chlordane	ND to 1000 ng	70. ng
	Chlorpyrifos	ND to 10,000 ng	150. ng
	4,4'-DDD	10-100 ng	5. ng
	4,4'-DDE	5-1000 ng	5. ng
	4,4'-DDT	10-4500 ng	5. ng
	Dieldrin	50-1000 ng	5. ng
	Heptachlor	2-500 ng	1. ng
	Malathion	ND to 1000 ng	110. ng

Water -- Metals: The water sample will be taken from the primary drinking and cooking water supply; if these are different, half the sample will be taken from each. If tap water, the system will be flushed for two minutes; if bottled water, the system will not be flushed. The pH of the water will be measured before collection. A sample of 1000 mL will be taken into a plastic container cleaned as per EPA *Handbook for Sampling and Sample Preservation of Water and Wastewater*. The sample will be sealed and chilled to #4EC. If the pH is below 5 or above 9, shipping to the laboratory will be expedited so that the sample can be extracted or have its pH adjusted within 72 hours of collection. Analysis will be done by EPA-EMSL Cincinnati in accordance with EPA Method 200.8. (See SOP F07.)

Expected concentrations and limits of detection: see Table 9.

Table 9 -- Expected Concentrations and LODs for Water and Food Sampling

Category	Pollutant	Drinking & Cooking Water		Food & Beverages	
		Expected Concentration	Expected Limit of Detection*	Expected Concentration	Expected Limit of Detection*
Metals	Arsenic	< 10. ng/g	1.0 ng/g	3.96 ng/g	3 ng/g
	Cadmium	< 5.0 ng/g	0.05 ng/g	20.1 ng/g	2 ng/g
	Chromium	< 50. ng/g	0.10 ng/g	1-2 ng/g	20 ng/g
	Lead	< 50. ng/g	0.7 ng/g	23.7 ng/g	10 ng/g
Pesticides	Chlordane		1.5 ng/g	0.023 ng/g	1.5 ng/g (cis)
	Chlorpyrifos		3 ng/g	0.111 ng/g	3 ng/g
	Chlorpyrifos-methyl		5 ng/g	0.295 ng/g	1-5 ng/g
	4,4'-DDD		5 ng/g		1-5 ng/g
	4,4'-DDE		2 ng/g		2 ng/g
	4,4'-DDT		3 ng/g	0.302 ng/g (total DDT)	3 ng/g
	Dieldrin		1.5 ng/g	0.076 ng/g	1.5 ng/g
	Heptachlor		1.5 ng/g	0.020 ng/g	1.5 ng/g
	Malathion		3 ng/g	1.781 ng/g	3 ng/g
PAHs	Anthracene				
	Benzo(a)-pyrene				
	Chrysene				
	Phenanthrene				

*Expected Limit of Detection Given Sampling Protocol

Water -- Pesticides: The sample will be taken as for metals. The container will be a glass bottle with Teflon-lined cap, supplied by Southwest Research Institute, cleaned as per EPA method 525.2. The pH will be measured and a 1000-mL sample taken as for metals. Analysis will be done by EPA-EMSL Cincinnati in accordance with EPA Method 525.2. (See SOPs F07, L13, and L14.)

Expected concentrations and limits of detection: see Table 9.

Food

Sampling for food contamination will be done in two ways. The first of these will be an in-home, four-day duplicate diet investigation. The target individual will be asked to prepare a duplicate plate of food and beverages as prepared for each meal. All foods collected in this manner will be stored in a small refrigerator. Respondents will be compensated $15 per day for food. Different people"s samples will be analyzed separately. (See SOP F08.)

In addition to the duplicate plate sample, the target individual will be asked to fill out a checklist describing the foods eaten during the four-day sampling period. The checklists will be analyzed to produce a normalized diet for the given stratum and then a shopping list representing foods eaten in quantities of at least " serving during the four-day period. Foods will be purchased at three neighborhood stores: a large supermarket, an intermediate-size store, and a small grocery store. The foods purchased will be used as a mini-market basket of foods commonly eaten by respondents. A single set of food items will be composited for each stratum (urban, suburban, rural) and store size within the 50-home sample. Thus for each Cycle, nine food samples will be obtained. (See SOP F09.)

We will then test the hypothesis that a mini-market basket sample and a duplicate diet sample will produce identical results for the pollutants studied. If this is true, the burden on respondents associated with diet-related exposures can be greatly reduced, requiring only the checklist.

The details of the food analysis procedures including storage, preparation, and shipping procedures are in SOP L15 and the FDA SOPs.

Expected concentrations and limits of detection: see Table 9.

Biological Samples

The primary goal of this study is to evaluate the utility of certain short-term exposure assessment techniques to predict long-term exposure. Urine and blood concentrations will be used as a measure with which to work. The environmental and questionnaire data will be compared with these biological markers.

The urine samples collected will be first morning void, taken on days 2 and 8. Although not as truly representative of exposures as a 24-hour sample would be, the first morning void carries less respondent burden. It also is more representative of exposure than the more-convenient spot urine sample. Sampling will use a sterile plastic collection cup with a lid. The day 2 sample will be analyzed for metals, the day 8 sample for pesticides and PAHs. (See SOP F10 in Appendix).

Blood samples will be taken on day 2 via venous puncture using a Vacutainer collection device, by a phlebotomist licensed in Maryland. Because of the universal interest in these samples, not only for metal analysis but for other pollutant classes as well, seven tubes totaling 56 cc will be taken from a single puncture. (See SOP F11 in Appendix.)

All target individuals will agree in advance to provide all samples including blood samples; potential target individuals who do not agree will not be selected. Because of the difficulties of blood sampling including the increased burden placed upon the target individual, our protocol allows for item non-response for blood samples after initial completion of some cycles. That is, if the target individual changes his or her mind about blood sampling after the study is already under way, the other samples from that household will still be taken rather than waste the cost of initial sampling.

Analysis will be done by the CDC according to standard protocols.

Expected concentrations and limits of detection: see Table 10. Limits of detection remain to be determined. Collaborating agency will address this issue with Harvard and NIST.

Table 10 -- Expected Concentrations and LODs for Urine and Blood Sampling

Category	Pollutant	Urine		Blood	
		Expected Concen-tration	Expected Limit of Detection Given Sampling Protocol	Expected Concen-tration	Expected Limit of Detection Given Sampling Protocol
Metals	Arsenic		2.0 : g/dL		
	Cadmium		0.2 : g/dL		0.5 : g/dL
	Chromium		2.0 : g/dL		
	Lead			1-50 : g/dL	0.5 : g/dL
Pesticides	Chlordane				0.5 : g/dL
	Chlorpyrifos		0.5 : g/dL		
	4,4'-DDD				
	4,4'-DDE				
	4,4'-DDT				
	Dieldrin				
	Heptachlor				
	Malathion		0.5 : g/dL		
PAHs	Anthracene				
	Benzo(a)pyrene				
	Chrysene				
	Phenanthrene				

2.1.4 Respondent Activities

During the course of this investigation, we will be asking our respondents to perform several activities. Among these are the maintenance of records for: (a) a four-day daily food checklist, and (b) a seven-day daily activity and location diary (covering days 2-7 and parts of days 1 and 8).

The data collected on environmental variables are necessary to evaluate the influence of all media on individuals' exposures. Because the study is designed to assess the ability of short-term measurements to predict long-term exposures, it will be necessary to return to each respondent's residence repeatedly. Each of the eight Cycles requires three visits to each home.

Our protocol is designed to result in minimum burden on the respondent. This is especially necessary in the Harvard group study as we plan on returning numerous times to the same location. None of the environmental samples require work by the respondent, except the personal air sample, which requires the target individual to wear a sampler for 24 hours, and the duplicate food sample, which requires collecting food. Dust and soil samples, and a dermal wipe, will be collected at the first visit (day 1) of each Cycle. The air samples require pumps and samplers in the activity room and

outdoors for the seven-day sampling period, which places a burden on the respondent. The second visit (day 2) will be by a phlebotomist who will take a blood sample and collect a urine sample and the personal air sampler. The third visit (day 8) will be by a technician who will remove the air sampling apparatus; and the interviewer who will administer the followup questionnaire, examine the diaries, collect the food and a urine sample, and take a second dermal wipe.

Between visits, the individual will be asked to wear the personal air monitor from day 1 to day 2, give urine samples on the mornings of days 2 and 8, and collect food on days 3-6.

2.1.5 Sample Collection Time Estimates

Tables 11-13 present time estimates for the three visits of a Cycle (days 1, 2, and 8). Estimates are based on dress rehearsals conducted at the homes of Harvard employees. Each vertical box equals 10 minutes; m = minutes; Q = questionnaire.

Table 11 -- Sampling Schedule, Visit 1 (Day 1)

Time (minutes)	Interviewer	Field Technician 1	Field Technician 2
0-10	Introduce staff to respondent. Ask about activity room, yard, etc.	Be introduced. Look for good places for air samplers.	Be introduced. Look for good places for dust sampling.
10-20	Begin Baseline Q (40 m).	Set up outdoor air sampler (40 m).	Begin floor plan (25 m).
20-30	Continue Baseline Q.	Continue outdoor air.	Continue floor plan.
30-40	Continue Baseline Q.	Continue outdoor air.	Finish floor plan.
40-50	Continue Baseline Q.	Finish outdoor air.	Begin dust (30 m), starting in activity room.
50-60 (1 hour)	Activity diary instructions (5 m)	Start indoor air (40 m).	Continue dust.
60-70	Duplicate food instructions (10 m).	Continue indoor air.	Finish dust; pack tools.
70-80	Instructions: Food checklist (5 m). Urine sample (5 m).	Continue indoor air.	Technician Q (8 m).
80-90	Dermal wipe (8 m). Technician Q (floor cleaning) (3 m).	Finish indoor air.	Begin soil, including yard and foundation plans (40 m).
90-100	Water (5 m). Begin personal air instruction (15 m).	Prepare personal air sampler (5 m) Begin personal air instruction (15 m).	Continue soil.
100-110	Finish personal air.	Finish personal air.	Continue soil.
110-120 (2 hours)	Thank respondent, remind him/her of days 2 & 8 appointments.	Pack tools.	Finish soil; pack tools.

Table 12 -- Sampling Schedule, Visit 2 (Day 2)

Time (minutes)	Phlebotomist
0-10	Take blood samples
10-20	Pick up urine sample, reminder for urine sample 2
20-30	Collect personal air sampler; take flow reading

Table 13 -- Sampling Schedule, Visit 3 (Day 8)

Time (minutes)	Interviewer	Field Technician 1
0-10	Greetings, description of what we will do today (5 m). Check activity diary (5 m)	Begin taking down outdoor air samplers (25 m).
10-20	Begin followup Q (20 m).	Continue outdoor air.
20-30	Finish followup Q.	Finish outdoor air.
30-40	Dermal wipe (8 m). Collect duplicate food (7 m).	Begin taking down indoor air samplers (25 m).
40-50	Check food checklist (5 m).	Continue indoor air.
50-60 (1 hour)	Food diary followup Q (10 m).	Finish indoor air.
60-70	Collect urine sample (5 m). Thank respondent, confirm dates for next Cycle (5 m).	Pack equipment.

2.1.6 Sample Management

Sample Handling: The handling of samples will be carried out with the aim of minimizing possible loss and/or contamination of the sample. Steps to further this goal include the appropriate selection and acceptance testing of plastic sampling and storage containers, as well as of sampling devices, e.g., plastic syringes for the soil sample collection. See Section 5 and the SOPs in the Appendix, particularly SOP G05 "Storage and Shipping of Samples."

Sample Custody: An established procedure for the handling, storage, and transport of samples is necessary in order to eliminate sample contamination, loss, or tampering. In addition, the data pertaining to a sample must be traceable. Hence, a "paper trail" is created for each sample. A unique ID number will be assigned to each sample (see SOP G03 "Identification Numbers for Samples and Forms." Printed labels will show the ID number in bar-code and human-readable format. Identical labels will be affixed to the sample container, the logsheet, and the chain-of-custody form.

The chain-of-custody form (see SOP G04 "Chain-of-Custody and Sample Tracking") will accompany the sample wherever it goes. Anyone who receives, transfers, or ships the sample will sign and date the form and keep a copy. This information and a database management system will ensure knowledge of the status and location of any sample at any time, including retrospectively. All data entries are dated and initialed by the field or laboratory technician. Corrections are dated and initialed, while ensuring that the original entry information is preserved. Photocopies of sample logs are made and kept in a separate location from the original.

Sample Preparation: Sample preparation is discussed in the appropriate SOP for each medium.

Sample Analysis: Sample analysis is discussed in the appropriate SOP for each medium.

After samples are analyzed, remaining extracts will be stored until the data have been satisfactorily analyzed. Samples will remain in storage for at least one year or as long as their integrity can be maintained. Prior to disposing of any sample extract, the data will first be examined and analyzed, e.g. for outliers or any apparent discrepancies. In addition, it will be ensured that the sample analysis was carried out with all QA/QC steps followed (e.g., with regard to the calibration curve R^2, the

concentration of the standard reference material analyzed, etc.).

2.1.7 Data Analysis

The data analysis phase of this investigation is designed to address the temporal variability issue. Two forms of analysis will be done. First, data collected in each individual Cycle will be analyzed to determine distributional characteristics. This analysis will afford a description of the respondent's exposure, stratified by geographic area, for a single time period and thus can be viewed as similar to those data collected in other NHEXAS investigations. (See Appendix 3, Statistical Issues Related to NHEXAS Design, for a discussion of the effects of stratification.)

For analysis purposes, the data are then considered to consist of eight separate sub-datasets. For each sub-dataset, parametric and non-parametric statistics will be obtained that describe the data. Test of normality will be made prior to analyses relying on this assumption. If the data are not normal, simple transformations (e.g., log transformation) will be invoked and normality tested again. If the data are still found to be non-normally distributed, non-parametric tests will be used to back up all tests relying on the normality assumption.

Analysis will then proceed by comparing these statistics across the sub-datasets representing different Cycles. Statistical tests such as multiple t-tests and analyses of variance will be invoked to test whether the distributions differ. If all distributions are identical, then selection of any short-term monitoring period will be sufficient to measure the annual average exposure. Any differences in distributions will call the null hypothesis into question and suggest that multiple monitoring time periods may be needed. Caution will be exercised to apply rigid criteria for rejection of null hypotheses.

Prior to making such a decision, a further analysis will be undertaken. In a sense, each respondent will act as his or her own control. Analysis can thus proceed considering each respondent separately, resulting in 50 sub-datasets consisting of eight measurements each. Analysis on these 50 sub-datasets can proceed in a manner analogous to that outlined above, with the distributions now representing the variability on measurements over a one-year period. Statistics such as the mean and standard deviation for each respondent can be constructed. Analysis can then proceed by examining the magnitude of this variability relative to the mean value using, for example, the coefficient of variation in this temporal sense.

The next level of analysis will proceed by evaluating which component of variance is larger, that is, whether the variance associated with the population variability in a given Cycle is larger or smaller than the variance associated with temporal variability. If the population variability within a given Cycle dominates, then temporal variability is small and may be of lesser importance. If temporal variability is larger than or comparable to population variability, multiple period monitoring will be necessary for the full-scale NHEXAS project. This analysis can be accomplished by two-way ANOVA with both a within-Cycle and between-Cycle component.

2.2 Data Quality Indicators

2.2.1 Accuracy Requirements

The accuracy of an analytical technique reflects the extent to which there is an agreement between a sample's known concentration and the actual measured value. The accuracy in determining the pollutant concentration will be assessed by comparison with certified standards or special Standard Reference Materials (SRMs) under development for use in NHEXAS investigations by the National

Institute of Standards and Technology (NIST) at Gaithersburg, Maryland. NIST will further provide a major service by acting in a quality assurance coordination role for all NHEXAS studies.

Table 14 shows typical expected values for accuracy and precision (section 2.2.2).

Table 14 -- Accuracy and Precision

Medium	Metals		Pesticides		PAHs	
	Accuracy	Precision	Accuracy	Precision	Accuracy	Precision
air	"10%	"10%	"40%	"20%	"30%	"20%
water	"10%	"10%	"30%	"20%	"30%	"20%
soil & dust	"10%	"10%	"30%	"20%	"30%	"20%
dermal wipe	"10%	"10%	"40%	"40%	"30%	"30%
food	"20%	"20%	"40%	"15%	"40%	"15%
urine	"15%	"15%	"30%	"15%	"30%	"15%
blood	"20%	"20%	"40%	"20%	"40%	"20%

Assessment of accuracy will be made by calculating the percent recovery (%R) of known reference materials with an appropriate matrix. The percent recovery is calculated as follows:

$$\%R = (\text{measured value} / \text{known reference value}) \times 100$$

Although Table 14 refers to classes of compounds, e.g., pesticides, any individual compound within the class may vary slightly from the class value in the table. For example, for malathion in food, the recovery may be less than 75%. Table 15 gives expected recoveries.

Table 15 -- Expected Recoveries

Medium	Metals	Pesticides	PAHs
air	"10%	"20%	"20%
water	"10%	"20%	"20%
soil & dust	"15%	"20%	"20%
dermal wipe	"10%	"40%	"30%
food	"20%	"25%	"25%
urine	"20%	"20%	"20%
blood	"20%	"25%	"25%

If the %R value for a reference compound analyzed as part of a set of samples does not meet the acceptance criterion, then the analysis of the samples in that batch is suspect and they may need to be re-analyzed. If re-analysis is not possible, such data will be flagged.

2.2.2 Precision Requirements

The precision of sample measurement and the analytical technique will be assessed by collecting duplicate samples and calculating the relative standard deviation (RSD%). The RSD% is calculated for each set of duplicates:

$$RSD\% = [SD / mean] \times 100$$

where

$$SD = \sqrt{\frac{\sum_{i=1}^{N} (x_{i\text{-}first} - m_i)^2}{N - 1}}$$

and

$$mean = \frac{\sum_{i}^{N} \frac{(x_{i\text{-}first} + x_{i\text{-}second})}{2}}{N}$$

where $x_{i\text{-}first}$ = value of first duplicate measurement

$x_{i\text{-second}}$ = value of second duplicate measurement
m_i = mean of the two duplicate measurements
N = number of duplicate sets

The RSD% will be examined for each set of duplicates. Generally, samples with RSD% $ 15% are unacceptable. Such samples will be re-analyzed until the RSD% is acceptable; usually one re-analysis is sufficient. RSD% is a function of the absolute concentration. Samples with low concentrations will have measurement error at or near the measurement itself. Thus, poorer relative precision is expected for samples with very low concentrations. Hence, the $ 15% rejection criterion will not apply to blank samples and very low-concentration samples. For such samples, an absolute error criterion will be used to assess precision. For example, if instrumentation is available that is, itself, precise only to $\pm$ 1 ppb, then a reading of a sample of nominal 1 ppb concentration will not be rejected if it reads 0 ppb or 2 ppb. Very specific criteria will be developed in each laboratory for very low level concentrations. We will use diluted SRMs from NIST to aid in this assessment. See specific SOPs for details.

2.2.3 Comparability Requirements

Comparability of the data collected and analyzed over the course of the study will be ensured by a combination of training, audits and comparison studies. All field technicians will be trained by the same team of study coordinators from Emory and Harvard and will be given identical copies of the Field Technician Manual"; they will be instructed how to use the Manual as well as having it available as a reference.

Field audits will be conducted periodically by the Quality Assurance Officer and/or other QA staff to assess the performance and standardization of data collection practices both within and between sets of field technicians. Comparability of analytical chemistry results will be evaluated by routinely conducting analyses of certified standard reference materials, certified standard solutions and mock field samples. In addition, we will endeavor to establish an inter-laboratory comparison program in which either duplicate samples or selected split samples will be shipped to different labs for analysis; the results will be subsequently evaluated and corrective actions made when there is a discrepancy. Comparability will also be evaluated by examining results from blank and replicate samples over time within and between sets of field technicians using statistical trend tests and analysis-of-variance techniques.

Data SRMs should always be within 5% of their target. In some cases (e. g., mock standards) data should be within 10%of their target. There should be no discernible statistically significant difference in inter- or intra- field team derived data at the 95% confidence interval.

2.2.4 Completeness Requirements

Completeness for a pollutant measured in the study is defined as the percentage of valid samples obtained from those scheduled. The degree of completeness will be assessed after running a given number of samples. Completeness is calculated using the following formula:

% Completeness for parameter Y = (# of valid samples / # of total samples) H 100

The objective is a completeness level of at least 90% for each parameter. If completeness is less than 90% for any parameter, this will be immediately brought to the attention of the Principal Investigator. Appropriate corrective action will be considered and implemented.

2.2.5 Representativeness Requirements

The environmental sampling protocol will be written in such a way that the samples collected best reflect the soil, dust, water, and other media with which an respondent comes in contact.

2.2.6 Duplicate Requirements

For each of the matrices requiring duplicate samples a total of 10% of the total number of samples will be obtained, of which 1%-3% of the total number of samples will be sent to an external lab for analysis as part of the QA program. This will include air samples, soil samples, water samples, and dust samples (if available). These QA samples will be randomly selected from the pool of QA samples. The external lab will be expected to use their routine SOP. The agreement to be expected between the EUSPH lab(s) and the external QA lab(s) will be determined by 1. the analyte, 2. the matrix, and 3. the concentration of the analyte in the final solution. For example, at higher levels of the calibration curve, duplicate air samples for metals should agree to within 20% of the average of the two laboratories whereas at levels approaching the LOD, an agreement within 50% would be considered acceptable.

If results of the duplicate analyses fall out of the prescribed or expected range, an investigation will be initiated with the problem reporting SOP until the problem has been identified, corrected, and appropriate corrective actions instituted to prevent recurrence. All aspects of this will be documented in a detailed final report which will be distributed to concerned parties as well as to be appended to the duplicate analysis report that will be filed. Any revisions to SOPs will be made and the date recorded as well as noting within the SOP what was changed and why.

Determination of precision will be done as described in Section 2.2.2, Precision Requirements. Table 14 summarizes duplicate and replicate protocols. See the SOPs for more detail.

Table 16 -- Duplicates and Replicates

Sample		Number Collected*	QA Samples (Duplicate)	Replicate Sample Aliquots	Replicate Analyses	Blanks		Spikes
Medium	Analyte					Lab	Field	
Air, indoor	metals	1	10%**	none	10%	5%	10%	5%
	pesticides/PAHs	1	10%	none	10%	5%	10%	
Air, outdoor	metals	1	10%**	none	10%	5%	10%	5%
Air, personal	metals	1	10% wear two PEMs	none	10%	5%	10%	5%
Dust	metals, pesticides/ PAHs	1 (split for extraction)	10% on adjacent floor area	10% if quantity sufficient	10%	5%	10%	5%
Soil	metals, pesticides/ PAHs	1 (split for extraction)	10% duplicate cores	10% split & analyze composite	10%	5%	10%	5%
Dermal wipe	metals	1	none	none	10%	5%	10%	5%
	pesticides	1	none	none	10%	5%	10%	5%
Water	metals	1	10%	none	10%	10%	10%	
	pesticides/ PAHs	1	10%	none	10%	10%	10%	
Food	metals, pesticides/ PAHs	1 (split for extraction)	none	10% in 1 stratum per Cycle		none	none	
Urine	metals, pesticides	2	none	10% split		***	none	
Blood	metals	2 tubes	none	10%		***	none	
	pesticides	1 tube	none	10%		***	none	
	PAHs	2 tubes	none	10%		***	none	
	VOCs	2 tubes	none	10%		***	none	

*Number collected per household per Cycle.
**Indoor air: duplicate metal and pesticide/PAH samples will be taken at the same time, using one pump, to reduce burden.
***At discretion of analyzing laboratory.

2.3 Project Responsibilities

This work reflects a Cooperative Agreement between the United States Environmental Protection agency and the various institutions in the Harvard/Johns Hopkins Consortium.

Key personnel include:
P. Barry Ryan, PhD: Professor, Rollins School of Public Health of Emory University (EU-RSPH) and Adjunct Associate Professor of Environmental Health, Harvard School of Public Health (HSPH). Principal Investigator (PI).

Dr. Ryan will have oversight of all aspects of the Harvard/Emory/Johns Hopkins/Westat NHEXAS-related work. He is principally responsible for the design and implementation of the study. He will oversee the laboratory work done at Rollins School of Public Health of Emory University and

will act as liaison between the Harvard/Emory group and the cooperating agencies.

Karen A. Hammerstrom, B.S., J.D. Environmental Scientist, United States Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment, Exposure Assessment Group. Principal USEPA Collaborator.

Ms. Hammerstrom is the chief contact and collaborator between EPA and the Harvard/Emory group. She works as a collaborator on the design phase of the investigation. Additionally, she is responsible for liaison work between the Harvard collaborating and EPA. She will have major responsibilities in the implementation, data analysis and write-up of the project as a whole.

Thomas A. Burke, PhD: Associate Professor of Environmental Health, Johns Hopkins University (JHU). Co-Principal Investigator and Coordinator of full survey in Maryland.

Dr. Burke will have primary responsibilities in the daily oversight of activities in the greater Baltimore area. He will oversee the sample collection process and the flow of samples and data from Westat to Harvard and the cooperating agencies. In addition to these management responsibilities, Dr. Burke will act as Community Liaison and Public Contact Officer. In this capacity, he will meet on a regular basis with community groups, devise strategies for the proper dissemination of NHEXAS data to the public, and deal with state, county, and local officials in Maryland. Dr. Burke brings 15 years's experience in state and local public health to the project.

Halfk _zkaynak, PhD: Project Manager and Lecturer, HSPH. Manager of Preliminary Exposure Assessment Project.

Dr. _zkaynak is manager of the Preliminary Exposure Assessment Project under the Harvard/Johns Hopkins Cooperative Agreement. In this capacity his is chiefly responsible for the quality of the initial assessments in Region V and in Arizona. He also is involved in the design phase of the NHEXAS Pilot and will be responsible for data analysis and data management tasks.

John D. Spengler, PhD: Professor of Environmental Health, HSPH. Co-PI.

Dr. Spengler will bring 20 years of experience to the design phase of the NHEXAS Pilot investigation. He will act as an advisor and consulting scientist on the project.

Robert Weker, BS: Chief Analytical Chemist, Department of Environmental Health, HSPH. Quality Assurance Officer for NHEXAS Pilot Projects.

Mr. Weker will be responsible for all aspects of Quality Assurance in the Harvard/Johns Hopkins study. He is responsible for the development, validation, and implementation of all Standard Operating Procedures involving either field or laboratory work. He will be responsible for the quality of the data coming from the Harvard Trace Metals Laboratory. Additionally, Mr. Weker will offer advice in modification of protocols for field work and will act as liaison to the cooperating laboratories supplying analytical services to Harvard.

Robert Clickner, PhD: Westat. Manager of Field Study and Field Coordinator

Dr. Clickner will be responsible for all activities carried out by Westat Inc. In this capacity, he will act as field coordinator and will oversee all aspect of field data collection. He will manage the field teams and plan and coordinate their efforts. Dr. Clickner has also been instrumental in the design phase of the project. His statistical and field expertise will be used extensively throughout the investigation. Dr. Clickner brings 15 years of experience in statistical survey design to the project.

David Camann, MS: Director of Analytical Chemistry Laboratories, South West Research Institute. Manager of Pesticide Analysis.

Mr. Camann will be responsible for all activities at SwRI. In particular, he will oversee analytical activities associated with pesticide analysis as well as prepare of extracts for PAH analysis. Mr. Camann also has extensive experience in field operations for pesticide and PAH sampling. This expertise will be invaluable in final protocol development and modification.

Carol Botteron, MS: Research Specialist, HSPH. Harvard coordinator for all aspects of NHEXAS investigations.

Ms. Botteron is responsible for coordination efforts among the various collaborating groups, Harvard, Johns Hopkins, and the EPA. In addition, Ms. Botteron will be principally responsible for preparing draft reports for both EPA and the public.

David MacIntosh, ScD: Research Associate, Emory.

Dr. MacIntosh is responsible for components of Preliminary Exposure Assessment and is Project Data Coordinator. He will also take a leadership role in study management and data analysis at Emory University.

Robert J. Buck, PhD: Research Associate, HSPH.

Dr. Buck is responsible for components of statistical design and database management.

Bryan Burnette, MS: Research Specialist, Emory. Laboratory Supervisor, Emory Trace Metals Laboratory.

Lauren Sullivan, BS Administrative Assistant, EU-RSPH. Administrative Assistant for Emory Group.

Ms. Sullivan is responsible for administrative operation of the Harvard Group's activities.

In addition to these named individuals, other collaborating agencies play key roles in the Emory/Harvard/ Johns Hopkins NHEXAS study.

EPA/EMSL Cincinnati Laboratory: This laboratory will perform analyses on water samples collected in the study.

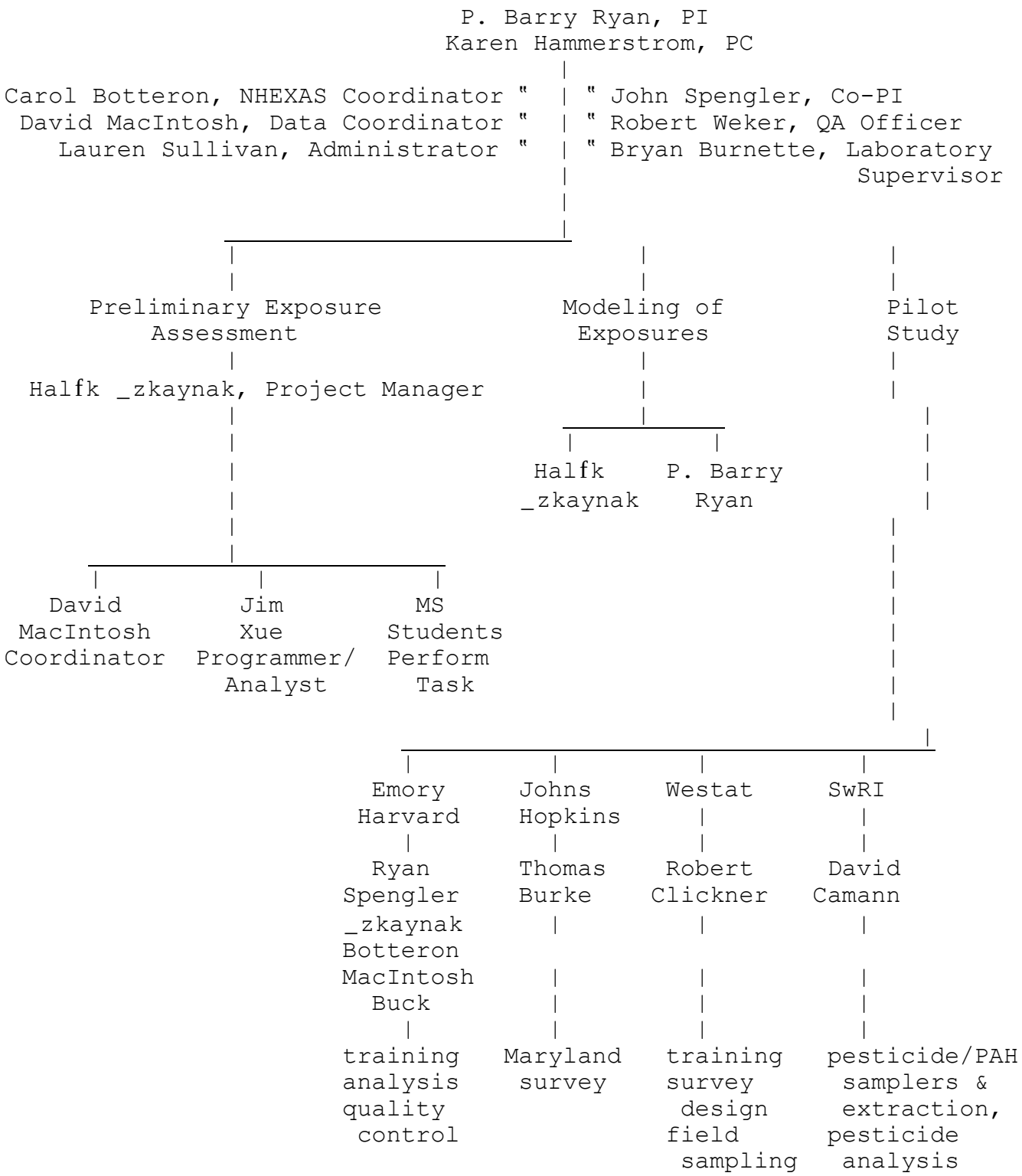
Food and Drug Administration Laboratory: This laboratory will analyze collected food samples for metals and pesticide residues.

Centers for Disease Control (CDC): The CDC laboratories will analyze biological samples (blood and urine) collected in this investigation for metals, pesticide residuals and metabolites, and other species. In addition, they will act in an advisory capacity to help the project in protocol development and sample collection. CDC will provide collection media for biological samples and give advice on shipping and handling procedures.

National Institute for Standards and Technology (NIST): NIST will play an extensive role in quality assurance procedures. They will act as the chief quality assurance laboratory. In addition, they will develop standard reference materials (SRMs) for the project and will set up quality assurance evaluations of the analytical procedures for other laboratories.

An organization chart is shown on the next page.

NHEXAS Pilot Study--Emory/Harvard--Organization



3.0 DATA ACQUISITION AND MANAGEMENT

3.1 Control and Calibration of Measurement and Testing Equipment

The sampling equipment used in the field requires control and calibration. This is most specifically true of the air sampling apparatus used to gather the indoor-outdoor air samples and the personal air samples. Depending upon the sampling device, calibration of these devices is effected by the use of either a transfer standard or an electronic calibration standard. The transfer standard used is normally a calibrated rotameter brought to the field to establish air flow via a linear calibration curve. The electronic calibration standard is used in conjunction with a previously-calibrated apparatus that uses a fixed electrical resistance to establish the flow rate. Comparison of the resistance with a digital voltmeter affords in-field calibration. The details of these procedures are found in the Standard Operating Protocols (see Appendix).

Laboratory measurement and testing instruments, including automatic dispensing pipettes and balances, are calibrated on a regularly scheduled basis and following the Standard Operating Protocols.

Analysis of metal samples will be accomplished through the use of atomic absorption spectroscopy (AAS-GF). Inductively coupled plasma"mass spectroscopy (ICP-MS) may be employed for confirmatory analysis. See SOPs L06 "Extraction of Metals from Sampling Media." Control and calibration of these instruments are addressed in SOPs L07 "Analysis of Metals by GF-AAS" and L08 "Analysis of Metals by ICP-MS."

3.2 Identification of Data

Each environmental sample collected will have a unique identification (ID) code that will be maintained throughout the life of the project. See SOP G03 "Identification Numbers for Samples and Forms." The code will identify parameters such as site, date, user, duplicate, etc. A detailed labeling scheme will be developed once the sample types have been finalized. Printed labels will show the ID number in bar-code and human-readable format. Identical labels will be affixed to the sample container, the logsheet, and the chain-of-custody form. Analysis on the sample will result in further data, which will be identified by the original cod.

Questionnaire data collected from an respondent will be maintained with a unique ID code for that respondent. Only selected members of the study team will have access to information on the respondent's identity. All information will be properly stored (as discussed in Section 4.4) beyond the end of the project lifetime. This information will be stored accompanied by any necessary explanations to ensure its proper usage during possible later examination.

3.3 Control of Erroneous Data

Logsheets will have a section for comments on each sample if they need to be made during the field collection and transport of environmental samples. The field staff will be instructed that these comments should cover points such as difficulties encountered during sample collection, any event that jeopardized the integrity of the sample, possible contamination sources, and so forth. On the laboratory or field logsheets, a staff member who decides that any datum must be rejected will cross it out by a single line, write a reason, date, and sign it. During data analysis, outlier analysis will be carried out to check whether certain data points need to be excluded from the analysis. If that is the case, a record of which points are excluded and the reason for doing so will be kept.

Missing value codes will be assigned to describe four different situations. The first three situations

apply to both physical samples and questionnaires: sample not collected, sample lost or ruined, and sample not collected because not applicable. For questionnaires "Don't Know" is a possible response and will be given a missing value code. Specific information on reasons for missing values will be contained in the relevant logsheets and logsheet database.

Electronic data input will be controlled in the following fashion. If data are obtained directly through computerized data acquisition, all data will be inspected by laboratory personnel to ascertain quality prior to finalizing datasets. Procedures will include, but will not be limited to, insertion of dummy quality assurance records to be handled identically to all other data and evaluated for consistency. The Quality Assurance Officer will do spot checks in addition to those routinely done by the data coordinator and data input personnel. See SOP D02 and SOP D03 for details on data entry.

For hand-entered data, in addition to the above, 100% of data will be entered by two individuals and compared electronically. This work will be done by a company that specializes in data entry. See SOP D04.

Laboratory results will be assessed to insure that data quality indicators are met. In addition, laboratory results above key "trigger" values (values which by regulation or inter-consortia agreement require reporting of results immediately to participants) will be flagged and mechanisms for such reporting implemented.

A sample tracking system will provide the location and status of any sample at any time including retrospectively. See SOP G04 "Chain of Custody and Sample Tracking."

3.4 Data Evaluation

 dard data evaluation procedures will include inspection of data in a univariate sense including development of distributional statistics (mean, median, etc.) and outlier evaluation. Prior to bivariate and multivariate data analysis, all data will be inspected via graphical procedures for appropriate bi- or multivariate outliers as well. Outliers will be flagged but not removed from the dataset. Outliers will be identified as being associated with equipment problems, known unusual circumstances (e.g., kitchen fires), or of unknown origin. The data analyst will be apprised of this through the presence of flag codes in the dataset. Analysis will proceed at the discretion of the analyst.

SOP D06 "Data Analysis," which is under development, will include protocols such as the identification and handling of outliers.

3.5 Procedures

Detailed Standard Operating Protocols are attached as an Appendix to this document. The titles are:

General

G01 Description of SOP Procedure
G02 Ensuring Confidentiality of Respondents' Records
G03 Identification Numbers for Samples and Forms
G04 Chain-of-Custody and Sample Tracking
G05 Storage and Shipping of Samples
G06 Problem Management
G07 Training of Field Technicians
G08 Training of Interviewers
G09 Training of Laboratory Technicians
G10 Training of Phlebotomists

G11 Training of FCC Staff

Field

F01 Field Sampling -- General Information
F02 Collection, Storage, and Shipment of Indoor and Outdoor Air Samples for Metal, Pesticide, and PAH Analysis
F03 Collection, Storage, and Shipment of Personal Air Samples for Metal Analysis
F04 Collection, Storage, and Shipment of House Dust Samples for Metal, Pesticide, and PAH Analysis
F05 Collection, Storage, and Shipment of Soil Samples for Metal, Pesticide, and PAH Analysis
F06 Collection, Storage, and Shipment of Dermal Wipe Samples for Metal and Pesticide Analysis
F07 Collection, Storage, and Shipment of Drinking or Tap Water Samples for Metal and Pesticide Analysis
F08 Collection, Storage, and Shipment of Duplicate Diet Samples for Metal, Pesticide, and PAH Analysis
F09 Administration and Analysis of Food Checklist and Purchase of Mini-Market Basket Food
F10 Collection, Storage, and Shipment of Urine Samples for Metal, Pesticide, and Creatinine Analysis
F11 Collection, Storage, and Shipment of Blood Samples for Metal, Pesticide, PAH, VOC, and Lipid Analysis
F12 Duplicate Sampling

Laboratory

L01 Purchase of Consumables
L02 Cleaning of Glass and Plastic Containers
L03 Operation of an Ultra-High-Purity Water System
L04 Balance Operation
L05 Sieving and Division of Dust and Soil Samples
L06 Extraction of Metals from Sampling Media
L07 Analysis of Metals by GF-AAS
L08 Analysis of Metals by ICP-MS
L09 Preparation of Exposure Media (PUFs and Quartz Fiber Filters) for Air Samplers for Pesticide and PAH Collection
L10 Extraction of Neutral Pesticides and PAHs from Air Sampling Media
L11 Extraction of Neutral Pesticides and PAHs from House Dust and Soil
L12 Extraction of Neutral Pesticides from Isopropanol Dermal Wipes
L13 Extraction of Neutral Pesticides from Drinking Water
L14 Determination of Pesticides and PAHs by GC/MS
L15 Preparation of Food and Beverages by Homogenization

Data

D01 Data Flow Procedures
D02 Questionnaire Data Entry and Database Preparation
D03 Lab Results Data Entry and Database Preparation
D04 Monitoring Data Flow
D05 Data Storage
D06 Exploratory Data Analysis and Summary Statistics

4.0 RECORDS USAGE AND MANAGEMENT

4.1 Data Records

This document outlines the data management procedures and storage of questionnaires and lab results. There is a large amount of data being collected and a complete list of samples collected are given in Tables 1-4. General categories of data collected are:

1. Questionnaires
2. Log sheets
3. Lab results
4. Chain-of-custody forms.

There are two main themes to the development of a complete data set which is readily accessible for data analysis. First, it is necessary to organize the collection and storage of data both in physical and electronic form. This section discusses only the storage of computerized data and paper forms generated during the study. Second, the data needs to stored in a way that allows easy manipulation of the data for the purposes of analysis and quality control.

A brief description of the process of creation of data records is given here. Details are given in Section 2 of the QSIP and in the Field, Lab and Data SOPs.

The interviewer will administer the questionnaires. The interviewer will:

- provide any clarification that the respondent asks for
- code any question that is not applicable or that the respondent refuses to answer
- make sure that the respondent understands how to complete the activity diary and food checklist
- ensure that all questionnaires, diaries, and checklists are complete upon collection

Information pertaining to the collection of environmental samples will be recorded on field logsheets by the Field Technician collecting the sample with any necessary comments being made on the logsheets. All such data will be dated and initialed. See the Field SOPs for details.

Information pertaining to the analysis of environmental samples will be recorded on computer files at the lab. Descriptive information such as date analyte determined, sample ID numbers, run number, batch number, etc. and data from the instrument, e.g., absorbances from spectrophotometers, will be recorded on computer disks.

The Chain-of-Custody forms will be used to monitor the progress of the samples and associated data values from the initial collection of samples to the entry into the official database.

4.2 Records Management System

A description of the records management system will be prepared and distributed to all individuals concerned with any aspect of data handling, be it generation, manipulation, or analysis of the data. The system will address issues including the following:

(1) general guidelines emphasizing the goal of generating quality data at all steps of the project
(2) description of the various data logsheets and logbooks used, and what is required in completing them (discussed in relevant SOPs)
(3) description of how to deal with erroneous data (discussed in Section 3.3)
(4) description of the steps involved in verification, identification, indexing, and retention of records (discussed in Section 4.4)
(5) guidelines on the transcribing of data from the various data sources to a central data filing system (discussed in Section 4.2.2)
(6) guidelines on the storage, preservation, safe-keeping, retrieval, and disposition of the original data

collection documents, as well as that of subsequently generated data records (discussed in Sections 2.1.6 and 4.5).

The data SOPs D01-D06 describe the record management system in more detail.

4.2.1 Description of Database System

The data will go through a series of transformations to be ready for thorough data analysis. Data preparation can be separated into four successive stages:

1. Initial data sets- The data as they come in from the field.
2. Complete Data Set (CDS) - the initial data has been collected into a single location and organized.
3. Analysis-ready Data Set (ADS) - the data are ready for analysis, but not stored in a form convenient for analysis.
4. Working data sets - the data are now in a form which can be used for analysis.

The first three stages will make up the database system. The Project Data Coordinator will oversee the maintenance of the database system and with the Principle Investigator decide who has access to the data. The Data Entry Supervisor and Data Entry Assistants will be responsible for performing the steps necessary to create the database system.

Initial data sets are the computerized version of the data sent from the lab or Data Entry Company and any paper forms from which the data originated. The initial data sets represent the official copy of original results stored on computer. The electronic reproductions may not be exact copies, but will not alter the data in any meaningful manner. Minor adjustments may be made (see below), but no summarization, deletion, *etc*. Data at this point will only be ready for the most cursory of statistical analyses. These data sets will be used to create the CDS and then archived.

The Complete Data Set (CDS) will contain all the original data and no information will be left out or summarized. This data set will mainly be used as an entry point to the original data when needed and as the starting point for constructing the ADS. The CDS will be considered the official database of the Harvard/Emory NHEXAS study.

The Analysis-ready Data Set (ADS) will be constructed with data analysis in mind. Several ADSs will be constructed because of the different types of data analysis that are of interest, such as response bias estimation or hypothesis testing. These data sets should be considered the baseline data sets for analysis. The ADS are general purpose data sets for analysis and are not intended to fulfill the needs of all research questions.

The CDS and ADS will be *Paradox* relational database systems stored on IBM-PC compatible computers. A complete discussion of the procedure for adding data to the CDS and ADS is given in SOP D02 and SOP D03. Depending on the interests of outside investigators requesting data, either the CDS or ADS will be the source for fulfilling these requests.

Working data sets are statistical or graphics software data files which contain the necessary data to answer the research question of interest. Typically they will be derived from the ADS, but for specialized questions the CDS may be used to generate these files. These data sets will be used to compute summary statistics, test hypotheses of interest and construct empirical distributions for the population studied. Given the large number of enquiries for this investigation, the creation of working data sets will be the responsibility of the investigators. Advice and assistance in creating working data sets will be given when required.

A note on nomenclature used in the rest of this document: *Tables* are data files that may be linked together in a relational database system.

Complete Data Set

The CDS will contain all data in an unchanged form. This is the baseline data set for the original data. There should be no need to search any prior data sets unless there is some concern that the data contained in the CDS are corrupted. The structure of the CDS can be outlined as follows:

1. A simple table containing participant ID, and list of cycles in which they participated will be the *master* table in the CDS.
2. Tables are constructed for each questionnaire and for each lab and contain all information from the relevant source.
3. Associated with each of these tables will be a *comment* table for comments on specific observations which have been noted outside the standard data collection format.
4. A table for information from logsheets which needs to be entered into electronic format.

Details on the structure of the CDS is given in SOPs D02, D03, and D04.

Small adjustments to the results from the questionnaires will be taken in the CDS.

1. Confidential questions will be placed in a separate table. This table will be encrypted for limited access and removed from the general data flow. Confidential questions include names, income levels, income levels, and other sensitive information.
2. Filling in the response to questions that were not asked, but whose results can be inferred from answers to other questions. See SOP D02 for a list of these questions and the inferred responses to be used.
3. Expanding questions with multiple responses into a series of binary response (yes/no, true/false, etc.) questions. For example, a question that asks to mark all that apply will need to be expanded so that each possible response is a variable with a yes/no response.
4. Several questionnaires, *e.g.* the descriptive questionnaire and technician walkthrough, will each be split into two tables because of the nature of information collected. For example, in the descriptive questionnaire, information collected on individuals living in the house will be contained in one table, which will be linked to the main descriptive questionnaire table.

Analysis-ready Data Sets

After the data has been collected into the CDS, it needs to be transformed in preparation for data analysis. These adjustments can be categorized into several basic types:

1. Summarizing the Food Diary Checklist, Time Diary and Activity Questionnaire. The unit of interest in this study is the participant-Cycle. For the Food Diary Checklist, Time Diary and Activity Questionnaire, data are collected on participant-Cycle-days. These data need to be summarized, typically by averaging, to get units of data to participant-Cycles.
2. Organizing Tables for Research Content. Data on a variety of subjects, such as food, activity, etc., are entered in a variety of locations. Broad categories are created and the data will be re-organized according to these categories.
3. Other summarizations and adjustments to questions. See SOP D02 for details.
4. Transformation of Lab sample data. The results from the sample need to be adjusted to reflect the measurement of interest. See SOP D03 for details.
5. Calculation of Total Exposure. Variables which sum the total exposure for a particular compound from different media or total exposure to all compounds for a particular media may be desired at this stage or may be deferred to the working data sets.

Four types of analysis will be carried out during the study: hypothesis testing, response bias estimation, and response rate estimation, and quality control testing. Only an ADS for hypothesis testing will be developed before the study begins, because of its central importance to the study. The participation summary data set will be used as the ADS for response rate estimation. The ADS will not be constructed

for quality control testing because working data sets can be constructed directly from the CDS. The response bias ADS will be constructed based on outside data which may be useful, such as census data.

Response bias estimation ADS: A response to the descriptive questionnaire will be requested from everyone contacted during the survey, regardless of whether they agree to participate in the study. This questionnaire, along with information on households which were not successfully contacted, will form the basis for the analysis of response bias. Other information which may be useful, if available, are: census information for the area and relevant information from other questionnaires which can be compared to the census data. The census data will be used as points of reference and may not need to be incorporated into a data set. The information contained in the questionnaires other than the descriptive questionnaire will only be useful (and hence needed) if census data is available and will be gathered into a single table with questions from the descriptive questionnaire.

Hypothesis Testing: The hypothesis-testing ADS will be the depository for all data to be used in all major investigations of the study. This is the baseline data set for analysis and will contain all transformations needed so data is ready for analysis. The structure of the ADS can be outlined as follows:

1. A simple table containing participant ID, and list of cycles in which they participated will be the *master* table in the CDS and the ADS.
2. Tables constructed from questionnaire results will be defined by question topic. Categories will be: Demographics, Activity, House characteristics (physical), House characteristics (living), and Food.
3. Tables separated by compound for the lab results. For example, tables will be constructed one for each compound, *e.g.,* lead, malathion, *etc.* Along with each result will be a flag identifying the validity of the result. Other lab results that are determined to be useful will also be included in these tables.
4. Tables separated by media for the lab results. For example, tables will be constructed one for each type of media, *e.g.* indoor air, soil, *etc.* The same additional variables besides exposure concentration which are included in the tables described in Point 3 will also be included here.

The reason for dividing the tables by variable classification is to try to reduce the number of tables in use at any one time. Lab results are repeated, but in tables which address two clearly defined goals: analyses based on compound type and analyses based on type of media. Since these two goals will be the basis of much of the analysis, it will be useful to have tables which are organized for these purposes prepared ahead of time. Details on how the ADS is created are given in SOP D02 and SOP D03.

4.2.2 Data Entry System

There are several different types of data, each of which follows a different route to entry into a computerized database. Each type of data follows a similar course: transformation from a physical sample to the initial data set; movement from the initial data set to the CDS; and creation of the ADS from the CDS. A brief description of how data moves from the field into the ADS follows. Further descriptions are given in D01, D02, and D03.

To help ensure an orderly addition of data to the CDS, data will be added in well defined blocks or *units* of data. The definition of a *unit* will depend on the type of data. Each set of results sent from a lab will be defined as a unit of that lab's data. All questionnaires collected from all individuals who are participating during a specified interval of time will be defined as a unit of questionnaire data. The labs and questionnaires are separated from each other because of the variability in when the data from

these sources will be available. By defining them as separate units they can be included as soon as they are available and will not have to wait for data from different sources.

The size of a unit of lab data will depend on the labs since they will determine the amount of data returned at any time. However, the size of a unit of questionnaire data will be controlled by the Project Data Coordinator. Number of subcycles will be used as the measure defining the number of participants included in a unit of data. This will give some consistency to the scheduling of data management. For example, a unit of questionnaire data may be all participants in the first two-week subcycle of the study. For the first two or three Cycles the unit of questionnaire data will be defined as a two week subcycle worth of questionnaires. After the first two or three Cycles the size of a unit of data may be increased to 3 subcycles (a Cycle) because the data collection methods will have been checked for irregularities and lab results will be returning, adding to the data management burden.

Handling of Initial Data Sets

Questionnaire responses will be collected on paper forms by the Field Interviewer and returned to the Field Coordination Center. After the completion of each subcycle all questionnaires for the completed subcycle will be photocopied at the Field Coordination Center and sent to the Project Data Coordinator. The Project Data Coordinator will ensure that the questionnaires are photocopied upon arrival and the photocopies sent to the selected Data Entry Company (DEC) for entry into a relational data base system based on the structure described in Section 4.2.1 and SOP D02. The originals will be stored in a secure, designated file cabinet. The data entry company will be one of the following companies:

Datasync
1224 Collier Rd.
Atlanta, GA 30318

Input Services, Inc.
1090 N. Chase Pkwy Suite 300
Marietta, GA 30067

M-S Data Systems
770 Old Roswell Rd. Suite I-100
Roswell, GA 30076

Access Data Inc.
2131 Kingston Court Suite 114
Marietta, GA 30067

National Business Systems
1430 W. Peachtree St. Suite 700
Atlanta, GA 30309

The data entry company will be required to follow specified data entry quality assurance techniques. The quality assurance techniques are described in SOP D02. Each data file returned to the NHEXAS project will be assigned a unique file name describing the contents and sequence of return. The Project Data Coordinator will receive the data files returned from the DEC and will be handed to the Data Entry Supervisor for addition to the CDS. The creation of the CDS is described below and in detail in SOP D02. Quality assurance procedures will be implemented to check the data sets received from the

DEC and also to check the creation of the CDS. Details on quality assurance procedures are given in SOP D02 and SOP D05.

Lab results from physical and biological samples will be sent on computer files to the Project Data Coordinator. The data files returned to the NHEXAS project will be assigned a unique file name describing the contents and sequence of return. All information received will be included in the CDS for lab data. Given the nature of the data there will be little opportunity for direct quality control checks by the Project Data Coordinator. The labs will be required to follow specified quality assurance procedures, see the relevant Lab SOPs, but indirect methods will be used to check for unusual observations. See SOP D03 and SOP D05 for further details on these methods.

Information from log sheets will be entered directly into a *Paradox* table. The data included are:

- Rotameter readings for air samples (Personal, Indoor/Outdoor Air Logsheets)
- Information from Sieving and Division of Dust Samples
- Area sampled and type of surface for Dust logsheet
- Soil characteristics from Soil logsheet
- All information from water logsheet

The Data Entry Supervisor will be responsible for entering the logsheet data and will use double entry techniques. See SOP D04 for details on transcription and quality assurance procedures.

Data Entry System for CDS

Upon receipt of computer files from either a lab or the data entry company, the following steps will be taken:

1. The file will be copied to the archive directory. The file will be renamed if necessary to reflect the NHEXAS file naming standards. See Section 4.4 and 4.5 for more details.
2. If the file is renamed the file will be copied onto the same disk as the original with the new name for use as cross-reference.
3. The disk with the original data will be stored in the designated location for archived disks.
4. The data will be imported to the relevant table of the CDS and the resulting file will be renamed to identify it as the latest version of that table of the CDS. The file will be stored in the CDS directory. Details on importing files are given in SOP D02 and SOP D03.
5. The new CDS will be copied to the CDS back-up disk.
6. The old version of the CDS will be removed from the CDS directory so that only the current version of the CDS is available.
7. The data will be checked for errors as discussed above and photocopies which are returned with the computer files will be destroyed as originals are on file.
8. A record of completion will be entered into the designated logbook.

The Project Data Coordinator will oversee these steps are taken. The Data Entry Supervisor and Data Entry Assistant will be responsible for performing these steps. These steps will be scheduled during one session to guarantee that all steps are carried out. See SOP D02 for details.

The initial data set and the CDS will be identical for the logsheet data. See **Handling of Initial Data Sets** for details.

Data Entry System for ADS

For many reasons it will be useful to create analysis-ready data sets before the end of the study. To ensure continued data integrity, the ADS will be constructed from the most recently constructed CDS. The creation of the ADS will be more time consuming than the CDS and will be updated after each Cycle has been completed. Before the completion of the first Cycle the CDS can be used to investigate the integrity of the data.

The transformations described briefly in Section 4.2.1 and in detail in SOP D02 and SOP D03 will

be written as *Paradox query* structures and saved for repeated use. The relational database system tables for the CDS and the queries needed to construct the ADS will be created and debugged before data collection. The Project Data Coordinator will be responsible for the construction of these tables and queries. Because the queries are preprogrammed the steps for creating the ADS are simple, but dependent on the ADS being created. The general procedure is:

1. Run the queries necessary to create the ADS.
2. Copy the ADS to the ADS back-up disk.
3. Remove intermediary files created during the creation of the ADS.

The creation of the ADS is dependent on intermediary files. If the Data Entry personnel fail to implement the queries correctly, a massive failure will result and the process can must restarted.

Data Flow Monitoring

As chain of custody forms are returned the current location of the sample will be entered into the custody tracking data set. The custody tracking data set will be contained in an Excel Workbook. See SOP D01 for more details on the custody tracking data set. The Data Entry Supervisor will be responsible for entry into the custody tracking data set.

Each day that involves starting a Cycle in a participants home (Visit 1) the FCC will e-mail to the Data Entry Supervisor a list of those households which at least initiated participation, *i.e.,* responded to the questionnaires or allowed samples to be taken. The Data Entry Supervisor will then enter into the Participation Summary Data Set the Participant ID, the Cycle number and the date the Cycle started for that participant. See SOP D01 for details on the Participation Summary Data Set. On Day 8 of the Cycle for a participant the FCC will e-mail the list of households participating in Day 8 of the Cycle. The Data Entry Supervisor will enter the date of Day 8 of the Cycle into the appropriate record of the Participation Summary Data Set. This will ensure correct entry of Participant ID and Cycle number into the Participation Summary Data Set.

4.3 Record Validation

All primary data recording documents (whether field logbooks, data sheets, laboratory notebooks, or computerized records) will be initialed or signed as well as dated. This authentication of the primary records will enable the tracing back of collected data to the individual collecting them.

Much data will be generated in this project, since numerous environmental sample concentrations and questionnaire data points will be generated for each respondent in the study. Thus, there will be a significant transcribing of data from primary data sheets to a computerized data records set. A thorough set of validation procedures will be needed for the entered data.

The basic approach will be the following:

(1) data are entered twice by different operators from the original documents; whenever the second entry is not identical to the first, the computer will alert the operator to resolve the discrepancy;
(2) field operator notes relating to any mislabeling or misidentification in the field are reconciled with the laboratory analyst's labeling and results;
(3) field and laboratory QC notes are reviewed and incorporated as needed;
(4) sample transmittal forms are updated and maintained;
(5) for final processing, any environmental sample with negative final concentration is investigated with respect to the blank correction that was applied.

4.4 Records Identification, Indexing, and Retention

Sample codes are discussed in Section 3.2.

Initial data files from labs and questionnaires may have a naming convention different from one used internally. The names of files sent to us will be compiled in a logbook as described in Section 4.2.2 as they enter the system and cross-referenced with the designated name using the NHEXAS file naming standard. Each sheet in the logbook will be set aside for a different type of data and the pages will be labeled accordingly. This record will include the name of the computer file returned, the name assigned to the returned computer file; the name of the updated CDS; the name of the person carrying out these tasks and the date that they were performed. The Data Entry Supervisor will be responsible for maintaining this logbook.

The CDS files will be named after the questionnaire or lab that describe the contents of the file. The ADS files can be named after the questionnaire or type of samples that describe the contents of the file. A numerical index will be attached as well to indicate the number of data units which have been added to file. A list of filenames is given in SOP D02 and SOP D03. The names of working data set files will be at the discretion of the investigator.

Field or variable names for questions in the CDS and ADS will be the question number when possible. Field names in other cases will depend on the information contained in the question. Variable names for lab data will contain information on the compound analyzed (if needed), the media sampled and the actual information in the variable. For example, the weight of the personal air monitor sample will be PA_WT. A glossary of field names will be compiled.

Upon completion of the data collection phase, the confidential file linking Participant ID and participant name will be purged from available computer systems and stored in a locked filing cabinet (along with all hard copies of records with name identifiers) for a minimum of five years. At that time, after consultation with the Environmental Protection Agency, the records may be destroyed.

4.5 Records Distribution and Storage

The original or primary data records will be delivered to the Data Input Supervisor. There will be limited access to the information enabling the decoding of a respondent's identification code. The computerized filing of data will be developed in such a way as to enable ease of content identification and information retrieval. An index of the data will be maintained and upgraded regularly.

Original data will be produced in two forms: written responses and physical samples. The storage of physical samples is discussed in SOP G05. Copies of questionnaires will be sent to the data entry company and the originals will be stored in a secure filing cabinet. Upon the return of the photocopies of the questionnaire they also will destroyed. Access to the filing cabinet will be limited to the Principal Investigator, the Project Data Coordinator, and the Data Entry Supervisor. The necessary information from logsheets and chain of custody forms will be entered into the CDS as described in Section 4.2 and the documents will be stored in the same manner as the questionnaires. The labs may send print outs of their results, but it is not expected. If they do, the printouts will be stored in the same manner as other written responses.

All data except confidential information will be stored in two locations on the designated IBM compatible computer, and on 3.5-in. floppy disks. Confidential information will only be stored on floppy disks. Storage of data on hard disk will be divided into three directories: the *archive* directory for the initial data files, the *CDS* directory for the complete data sets, and the *ADS* directory for the analysis ready data sets. Initial data files returned from the labs and the DEC will be stored in a secured area. A computer disk will be assigned for each category of tables for the CDS and ADS. Newly created versions of the CDS and ADS files will be copied to the appropriately designated disk and stored in a secured area. New volumes of these disks will be created as necessary. Access to the disks will be limited to the Principal Investigators, the Project Data Coordinator, and the Data Entry

Supervisor.

Confidentiality

Each respondent will be assigned an identification number that will be used in analysis. All data from questionnaires, activity diaries, etc, will be entered using this identifier. A single file will be maintained affording linking of the identifying number and the respondent. This file will be maintained, in encrypted form, for the duration of the data collection process. This is necessary to maintain records for contact of respondents.

Confidential information collected on the questionnaires will be entered into separate tables and will be encrypted upon entry into the CDS tables. The CDS tables will be removed from circulation by removing the CDS tables from the CDS directory and storing them only on disk. The initial data sets and CDS for confidential questions will be stored in a secured filing cabinet and access will be limited to the Principal Investigators and Project Data Coordinator. Information required to perform the duties of the study will be released on a need to know basis by the Project Data Coordinator.

The confidential portions of the questionnaire will be separated from the rest of the questionnaire and photocopied separately. A photocopy will be stored in a secure file cabinet at the FCC and only the Field Coordination Supervisor will have access. The originals will be sent to the Project Data Coordinator at Emory where the data will be entered as described above and in SOP D04. After data entry the original confidential questionnaire forms will be stored in a secure file cabinet at Emory. See SOP D01 and SOP D04 for details.

Data analysis will proceed with respondents identified only by identification number. No data will be released with names of respondents intact. All publication or dispersal of data will be accomplished through summary information in a manner designed to ensure that no individual records will be identifiable with a respondent. Other procedures may also be implemented in consultation with the EPA or other government entity as may be required.

5.0 ROUTINE CONTROLS AND PROCEDURES

5.1 Maintenance of Equipment

The atomic absorption spectrometer (AAS) will be used for the analysis of the metal concentration in the environmental samples. Maintenance occurs according to an established schedule. Logbooks are kept in the laboratory, keeping records of the time period each instrument was in use, the number of samples analyzed in that time period, and any problems encountered during the analysis. The graphite tube used in the AAS is conditioned every day prior to the start of analysis. See SOP L07 "Analysis of Metals by GF-AAS."

The micropipettes used for sample dilution are calibrated every month at the volumes typically used. A calendar of the calibrations and the calibration results are maintained in a special file. Balances and other laboratory instruments are maintained according to manufacturers' specifications.

Field equipment, particularly the air samplers, is maintained according to the SOPs.

5.2 Quality of Consumables

The reagents ordered, namely acids, are of reagent grade if used for cleaning and of trace-metal or super-analyzed grade if they are to be used in the analytical procedure. The date of container opening and expiration are recorded on each container. See SOP L01 "Purchase of Consumables."

If it is possible to analyze the reagents on the AAS for metal content (based on their chemical nature and concentration levels), this will be done prior to using the reagents in the analytical process.

The laboratory has a water purification system that produces Milli-Q water. This system tests the resistivity of the water, which should be within a determined range, as well as temperature. This system is routinely maintained as part of the laboratory's overall quality assurance program. For more details, see SOP L03 "Operation of an Ultra-High Purity Water System."

5.3 Labeling

Labeling of samples is discussed in Section 3.2.

The labels of all delivered supplies and materials will be checked to ensure delivery of the proper items (e.g., acids of the correct grade). Labels should include appropriate information relating to identification, composition, safety hazards, stability, storage and handling requirements. If a label is missing or thought to be incorrect, the supplier and/or manufacturer will be contacted and appropriate action will be taken. See SOP L01 "Purchase of Consumables."

5.4 Acceptance of Equipment and Materials

All glassware and sampling and storage containers are purchased with possible contamination from the material being an important consideration. Assessment will be made through analysis of blanks and development of quality control charts. Examination of blank values for indication of out-of-control processes will be assessed by the Quality Assurance Officer. SOPs F02 and F03 on indoor/outdoor and personal air describe acceptance testing of filters. All filters of a given type will be from the same lot.

Procurement documents will include specifications and other requirements that define the desired characteristics of the equipment or materials. Control over procurement documents will be established so that changes in specifications or other requirements are not made without proper review and approval. Acceptance of equipment and materials will be based on verification that specifications and other requirements have been met through inspection upon receipt or through supplier's certification. See SOP L01 "Purchase of Consumables."

5.5 Storage of Equipment and Materials

Established and rigorous cleaning procedures are in place for the various types of containers based on the material they are made of and their intended usage. See SOP L02 "Cleaning of Glass and Plastic Containers."

Upon washing, the sample collection and storage containers will be stored in clean plastic bags. All supplies and materials will be stored in the Trace Metals Laboratory, which has a well-controlled ventilation system to maintain temperature and humidity over a narrow range. In addition, storage of these items will be in a low-traffic area. There are no materials that need special storage with respect to limits on light exposure. All chemicals used for analysis (as opposed to cleaning) will be stored in appropriate cupboards in the laboratory's clean area.

6.0 TECHNICAL ASSESSMENT AND RESPONSE

6.1 Assessment Procedures

Every laboratory and field researcher in our consortium will maintain at least one laboratory notebook (or record notebook) and will review the contents of the recorded information on a weekly basis. Each researcher will document any technical problems or anomalies that are detected and also suggest the next steps to be taken to remedy the situation, as in SOP G06 "Problem Management." Instrumentation is expected to operate reliably. Any problems should be detected by staff or through duplicate analysis (Section 2.2.6). Equipment testing procedures are described in the SOPs. Questionnaires collected will be reviewed regularly to see whether any special difficulties develop, for example, whether certain questions are more often left unanswered compared to others.

As part of ongoing QA, internal audits will be conducted by Emory, HSPH, or JHU QA officers or staff. SOPs are being developed as procedures are finalized. Audits will be used to verify compliance with the designated procedures, emphasizing record-keeping and issues of accuracy and precision of sampling and analysis. Although many types of problems can be discovered through the routine procedures used every day, e.g., a standard curve is determined to be internally consistent with an acceptable correlation coefficient and y-intercept but with a slope 2x the usual, other problems can be more removed and elusive to identify. Researchers themselves are expected to critically evaluate data and results of experiments to discover if a problem exists. In addition there will be scheduled weekly meetings to discuss and evaluate the data so far obtained.

External audits (by EPA or an EPA contractor) will be performed both in the field and in the laboratories of the participating institutions. This is necessary to recognize and identify problems not detected by the performing group, e.g., sometimes those problems that go unnoticed due to accepting the unchanging repetitiveness of a procedure. External audits are necessary to supplement internal procedures; there will be an audit external audit of both lab and field in the second or third cycle of sampling and analysis.

6.2 Assessment Evaluation

The external audit team will carefully examine all systems, procedures, and personnel, then complete a written report to be distributed to all interested parties. When a potential or real problem is recognized, it must first be confirmed either by complete review of all the data and results and/or by repeating the experiment and achieving the same result. Once a real problem has been confirmed, either obvious corrective actions can be instituted or suggestions can be considered by the researchers, the PI, and the NHEXAS sponsors. All of these are documented in both formal reports of the problem and the estimated effectiveness of the proposed corrective actions.

6.3 Assessment Response and Follow-up

If any difficulties are associated with the environmental sampling protocol, the project manager and the field and laboratory personnel will discuss the problem and decide on any necessary modifications.

It is important to note, however, that modifications affecting the collection of data need to be minimized to the extent possible. This will ensure that the data collected through the lifetime of the project are consistent in that they are not measuring different variables at different times of the project.

For any problem areas that may have arisen and where modifications were made to the procedures, it will be of great importance to maintain an accurate record of the changes instituted and the dates at which such changes were introduced.

Following distribution of the Assessment Evaluation Report the proposed corrective actions are initiated and preliminary experimentation resumed, then resulting data are evaluated to determine whether or not the corrective actions in fact are working. This evaluation will be done by both the PI and the researcher and will be documented and distributed to all concerned colleagues. This report

and/or a final report on the subject will be sent to all concerned NHEXAS staff.

If changes to SOPs are necessary, EPA will be notified according to the amending protocols in Linda Porter"s March 9, 1995, NHEXAS Approval Process.

7.0 MANAGEMENT ASSESSMENT

7.1 Assessment Responsibility

The key element associated with project responsibilities is the communication among the groups (Emory University, Harvard University, Johns Hopkins University, Westat, and Southwest Research Institute). The contact persons for these groups are P. Barry Ryan (Emory and Harvard), Thomas Burke, Robert Clickner, and David Camann respectively. All are easily accessible by electronic mail via Internet. In addition, biweekly conference calls will be set up to make sure that management of the project is under control. The subject matter of these conference calls will be documented.

P. Barry Ryan, Principal Investigator for the study, will be responsible for the ultimate workings of all aspects of this project. Thomas Burke will be responsible for oversight of all activities in the Baltimore area. Robert Clickner will oversee the field activities associated with the investigation. David Camann will be responsible for sampling and analytical activities associated with the pesticide component.

Other key personnel include Halfk _zkaynak of HSPH, who will oversee the Preliminary Exposure Assessments; John Spengler of HSPH, who will act as senior advisor to the Project; Carol Botteron of HSPH, who will be Harvard coordinator for all aspects of NHEXAS investigations; David MacIntosh of Emory, who will be Data Coordinator, and Robert Weker of HSPH, who will be Quality Assurance Officer.

7.2 Assessment Types and Usage

Formal and informal systems and performance audits will be performed as appropriate. Details such as mechanisms for error searches are being planned.

7.3 Assessment Criteria

Reviewers and auditors will have technical expertise in the field, will not be directly involved with the work, and will have sufficient information about the work (including purposes and objectives) to adequately evaluate the work. Results of reviews and audits will be documented.

7.4 Assessment Documentation

Forms will be developed for corrective action and other purposes as appropriate. Results of reviews and audits will be documented. Reports will include date, place, participants, activities reviewed, evaluation process used, results of evaluation, and recommendations. These reports, including any notification letters and responses, will become part of the records associated with the project.

Reports will be sent to respondents describing results of the investigation. These will be of two forms. The first of these is an action report triggered by a specific pollutant-medium value in excess of a guideline developed either by EPA, State, or Local Agencies or by the Centers for Disease Control. These reports are sent to the participant immediately as required by law or regulation.

The second type of report informs the participants of the results of the investigation. These will not be sent to the participant until after the study is completed so that activities associated with exposure will not be modified by the content of such reports. These reports will give the individual his or her own exposure values as well as placing such values in the context of both the study population and potential health effects. In the past, the Harvard group has used a "gas gauge" representation of exposures with the "full" and "empty" marks corresponding to the maximum and minimum measured values respectively. The dial of the gauge represents the individual"s measurements. Such visual representations have been well-received.

8.0 TITLE AND APPROVAL SHEET -- see front of document
9.0 TABLE OF CONTENTS -- see front of document
10.0 DISTRIBUTION LIST

APPENDIXES

Appendix 1 NHEXAS Glossary

"	probability of false positive (Type I error)
A	in ID number: Geographic Stratum identifying letter. Valid values: U, S, R
AAS	atomic absorption spectroscopy
AAS-GF	atomic absorption spectroscopy with graphite furnace (GF-AAS)
ADD	average daily dose
ANOVA	Analysis of Variance
AREAL	Atmospheric Research and Exposure Assessment Laboratory (EPA)
As	arsenic
ATSDR	Agency for Toxic Substances and Disease Registry
AZ, UA	University of Arizona (Tucson)
$	probability of false negative (Type II error)
B(a)P, BaP	benzo(a)pyrene, $C_{20}H_{12}$, a toxic PAH in smoke
Battelle	Battelle Memorial Institute (Columbus, OH)
benzene	C_6H_6, a carcinogenic VOC found in gasoline and smoke
bioavailable	capable of being absorbed and available to interact with the metabolic processes of an organism
biomarker	biological marker of exposure, e.g., phenol is a biomarker for benzene
Black Box Sampler	air sampling device with timer and "flip-flop" apparatus allowing multiple inlets to be sampled
C	in ID number: Cycle identifying number. Valid values: 1# C # 8
C(t)	exposure concentration as a function of time
CAG	Cooperative Agreement (HSPH is a CAG for NHEXAS)
cc	cubic centimeter, equal to milliliter (mL)
Cd	cadmium
CDC	Centers for Disease Control (Atlanta)
cfm	cubic feet per minute
CFR	Code of Federal Regulations
ChE	cholinesterase (chlorpyrifos is a ChE inhibitor)
chlorpyrifos	an organophosphate pesticide, Dursban
%C	completeness: % of valid samples obtained from those attempted
CO	carbon monoxide
Cr	chromium
CT	contact time
Cycle	7-day period during which a household is studied
D_{app}	applied dose: amount of a chemical that reaches the exchange boundaries of the skin, lung, or gastrointestinal tract
D_{int}	internal dose (absorbed dose): amount penetrating across an absorption barrier or exchange boundary
D_{pot}	potential dose: potential amount that could be absorbed if it were 100% bioavailable
de minimis	insignificant (e.g., risk)
DQOs	data quality objectives: acceptable degree of uncertainty
E	magnitude of exposure (concentration H time)
ED	exposure duration

EMSL	Environmental Monitoring Systems Laboratory (EPA - Cincinnati)
EOHSI	Environmental and Occupational Health Sciences Institute (Piscataway, NJ)
EPA	Environmental Protection Agency (US)
FC	Field Coordinator
FCC	Field Coordination Center
FDA	Food and Drug Administration
Field Coordination Center	Westat in Rockville, MD. Location of preparation for field sampling, and storage and shipping of samples.
Field Interviewer	The NHEXAS Field Team's primary contact person, who is responsible for interaction with respondents; administers all questionnaires (except the Field Technician Questionnaire), gives instructions for duplicate diet sampling, and takes water and dermal wipe samples.
Field Technicians (FT1, FT2)	The NHEXAS Field Team members who are responsible for setup, maintenance, and breakdown of equipment brought to the field for sampling. FT1 is responsible for air sampling. FT2 is responsible for house, foundation, and yard plans, soil and dust sampling, and the Field Technician Questionnaire.
FTE	full time equivalent
GC	gas chromatography
GIS	Geographic Information System
GLPs	good laboratory practices (related to SOPs)
GPS	instrument to determine latitude & longitude
GF-AAS	atomic absorption spectroscopy with graphite furnace (AAS-GF)
GSD	geometric standard deviation
HANES	Health and Nutrition Examination Survey (3, first in early 1970s)
Harvard Black Box	see Black Box sampler
Harvard Impactor	device that removes particles larger than a certain size (e.g., 10 :m) from an air stream
HERL	Health Effects Research Laboratory (EPA)
HF	hydrofluoric acid
HI	[1] Harvard Impactor particle sampler [2] in ID number: household identification number
HNO_3	nitric acid
HPLC	high pressure liquid chromatography
HSPH	Harvard School of Public Health (Boston)
HVFS	High Volume Furniture and Surface Sampler (for dust)
HVS3	High Volume Small Surface Sampler (for dust)
I	in ID number: individual respondent identifying number
IAG	interagency agreement
ICP-MS	inductively coupled plasma--mass spectrometry
ICR	Information Collection Request (report for OMB)
ID	identification
IIT	Illinois Institute of Technology (Chicago)
IR	intake rate
isopropanol	isopropyl alcohol, rubbing alcohol
JHU	Johns Hopkins University (Baltimore)
l, L, R	liter
Laboratory	staff person at HSPH Trace Metals Laboratory or other laboratory who is

Technician	responsible for preparing and analyzing samples, and associated recordkeeping.
LOAEL	lowest observed adverse effect level
LOD	Limit of Detection
LOQ	Limit of Quantitation
LPM	liters per minute
matrix	a specific type of medium (e.g., drinking water) in which the analyte of interest may be contained
MEI	maximally (most) exposed individual, worst case
: g	microgram, 10^{-6} g = 10^{-3} mg = 10^{3} ng
: l, : L	microliter, 10^{-6} L = 10^{-3} mL
Monte Carlo analysis	repeated random sampling from the distribution of values for each of the parameters in a generic (exposure or dose) equation to derive an estimate of population distribution
MRL	minimal risk level
MSA	metropolitan statistical area
NAAQS	National Ambient Air Quality Standard
NHEXAS	National Human Exposure Assessment Survey (EPA)
ng	nanogram = 10^{-9} g = 10^{-6} mg = 10^{-3} : g
NIEHS	National Institute for Environmental and Health Statistics
NIOSH	National Institute for Occupational Safety and Health
NIST	National Institute of Standards and Technology
NOEL	no observable effect level
NOPES	Nonoccupational Pesticide Exposure Study (EPA, 1990)
NRC	National Research Council or National Response Center
OHR	Office of Health Research (EPA)
OMB	Office of Management and Budget
OMMSQA	Office of Modeling, Monitoring Systems, and Quality Assurance (EPA - DC)
ORD	Office of Research and Development (EPA)
PAH	polycyclic (polynuclear) aromatic hydrocarbon, e.g., benzo(a)pyrene
parameter	variable factor, e.g., site, date
Pb	lead
PCBs	polychlorinated biphenyls
PEM	personal exposure monitor (air particulates)
performance audit	obtains quantitative estimates of accuracy of measurement systems
phlebotomist	person licensed to take blood samples
PI	Principal Investigator
PIFS	Phase I Field Study (1994-1996)
$PM_{2.5}$	personal monitor sampling particulates # 2.5 : m
ppb	part per billion = ng/g = : g/kg
ppm	part per million = : g/g = mg/kg
PSU	primary sampling unit, e.g., a county
PTEAM	particle TEAM investigation (in California, for lead; see TEAM)
PUF	polyurethane foam, plug used to sample air for pesticides & PAHs
QA	quality assurance or quality assessment
QAPP	Quality Assurance Project Plan
QC	quality control
QMA	quartz microfiber filters, used in URG PUF cartridge

QSIP	Quality Systems and Implementation Plan
Qx	questionnaire(s)
R	in ID number: rural stratum (see A)
R	arbitrarily assigned integer used to identify one of multiple samples
%R	% recovery of a substance
RfC	reference concentration
RfD	reference dose
risk assessment	major areas: exposure assessment, hazard identification, dose-response assessment, risk characterization
RO	regional office (EPA)
RSD%	relative standard deviation percent, RSD% = [SD/mean] * 100
RSP	respirable suspended particulate matter
RTI	Research Triangle Institute (NC)
RTP	Research Triangle Park (NC)
S	in ID number: suburban stratum (see A)
SAB	Science Advisory Board (EPA)
SD	standard deviation
SMSA	standard metropolitan statistical area
SOP	Standard Operating Protocol (or Procedure)
SRI	see SwRI
SRM	standard reference material
SRS	simple random sample
ST	sample type identifying number (01 = baseline questionnaire, etc.)
stakeholder	anyone interested in study: respondents, government, industry, etc.
STEL	short-term exposure limit
SwRI	Southwest Research Institute (San Antonio, TX): extracts air samplers, dust, soil, and dermal wipes; and analyzes extracts for pesticides.
systems audit	evaluates operations and measurement systems of project, including training, facilities, recordkeeping, etc.
TEAM	Total Exposure Assessment Methodology studies (EPA, 1987)
TML	Trace Metals Laboratory
TSP	total suspended particulate matter
TWA	time-weighted average
t-test	tests difference of means in normal distribution
U	in ID number: urban stratum (see A)
UA, AZ	University of Arizona (Tucson)
URG	(company name) URG PUF cartridge used to sample air for pesticides & PAHs
Vacutainer	blood collection device
VOC	volatile organic compound, e.g., benzene, chloroform
V:V	volume:volume (dilution of acids etc.)
Westat	Westat, in Rockville, MD; does sampling
WHO	World Health Organization

Appendix 2 References

Dockery, D.W., Spengler, J.D., Reed, M.P., and Ware, J.H. Relationship among personal, indoor and outdoor NO_2 measurements. *Environ. Int.* **5**,101-107(1981).

Immerman, F.W., and Schaum, J. L. Nonoccupational Pesticide Exposure Study (NOPES). EPA Atmospheric Research and Exposure Assessment Laboratory, EPA/600/S3-90/003 Feb. 1990.

Kish, L. *Survey Sampling*. John Wiley & Sons, New York 1965**.**

MacIntosh DL, Xue J, Ozkaynak H. NHEXAS Pilot Exposure Assessment Study: Estimated Benzene Exposures and Adsorbed Doses in Region V and Arizona. Harvard School of Public Health. June 1994a.

MacIntosh DL, Xue J, Ozkaynak H. NHEXAS Pilot Exposure Assessment Study: Estimated Toluene Exposures and Adsorbed Doses in Region V and Arizona. Harvard School of Public Health. June 1994b.

Pellizzari E. Personal Communication to NHEXAS Working Group. March 1992.

Quackenboss, J.J., Spengler, J.D., Kanarek, M.S., Letz, R., and Duffy, C.P. Personal exposures to nitrogen dioxide: relationship to indoor/outdoor air quality and activity patterns. *Environmental Science and Technology* **20**, 775-783 (1986).

Schlesselman J. Planning a longitudinal study: II. Frequency of measurement and study duration. *J Chron Dis.* **26** 561-570 (1973).

Spengler, J.D.; Ferris, B.G., Jr.; Dockery, D.W.; and Speizer, F.E. Sulfur dioxide and nitrogen dioxide levels inside and outside homes and implications on health effects research. *Environmental Science and Technology* **13**,1726-1280(1979).

Spengler, JD, Schwab, M, Ryan, PB, Colome, SD, Wilson, AL, Billick, IH and Becker, E. Personal exposure to nitrogen dioxide in the Los Angeles Basin: Study design and results. *J. Air Waste Mgmt.* **44**, (39-47), 1994.

USEPA, Information Collection Request. NHEXAS Phase I Field Investigations. Submitted to Office of Management and Budget. April 17,1995.

Wallace, L.A., Pellizzari, E.D., Hartwell, T., Rosenzwieg, M., Erickson, M. Sparcino, C. and Zelon, H. Personal exposure to volatile organic compounds, I. Direct measurement in breathing-zone air, drinking water, food, and exhaled breath. *Environ. Res.* **35**,293-319(1984).

Appendix 3 Statistical Issues Related to NHEXAS Design

Introduction

This appendix discusses which data quality objectives can be met with the sample size chosen and an introduction to arguments about why those sample sizes meet the data quality objectives. In it we address several issues. First, we develop the general picture of the statistical needs of the study. We follow this by the a study of the chosen sample size with respect to quality of prediction of univariate distribution parameters. A discussion of the effect of attrition on these estimates follows

Project Goals

From a statistical point of view, several goals for the project may be envisioned**:**

1. Estimate specific population parameters and distribution for exposure.
2. Estimate seasonal trend in exposure for individuals and individual variability of exposure over time.
3. Test to determine whether selected socioeconomic groups are exposed to different levels of chemical of interest or have different seasonal trends or daily variability.
4. Test to determine whether individuals have the same trend and variability of exposure over time.

The first two goals are the main priorities. Fulfilling these objectives would bring an adequate conclusion to the investigation. The third goal is associated with looking at information on varaious factors as obtained by questionnaire and correlating exposures experienced by subgroups. The fourth goal would require examination of trends experienced by each individual independently. It is unlikely that for the sample sizes chosen (See Below) we can reach all four of these goals. From a brief overview of the problem, arguments can be made for the first three goals. We do not believe we will have enough power for the last goal.

Sample Size Considerations

It is necessary to estimate the sample sizes needed to meet the data quality objectives. There are three questions which can help gauge what might be possible to achieve for the four goals stated.

1. How many people must be included in the initial sample to achieve the necessary number of field observations?
2. How many individuals will be necessary to achieve the data quality objectives for estimating population parameters?
3. How many times must individuals be sampled to meet the data quality objectives for estimation and testing of individual parameters of seasonal trend and daily variability.

These questions can be pursued independently but all are essential in determining whether the study will be successful, i.e. it will meet the data quality objectives.

Two sample sizes need to be determined to assess whether the study will meet the data quality objectives, the population sample size n and the number of observations taken on each individual m. Both play a role in the estimation of population parameters, although m is not as important as n. For the other goals the size of m is the dominant factor since n will be sufficiently large because of the requirements of the first goal. The difference between the first two goals and the last two goals is the

shift from estimation to testing. Although they are inherently linked together it is easier to perform the estimation and determine the resulting confidence interval than it is to claim you are going to test poorly estimated (high std. err.) parameters and have the results be significant.

Study of response rate probability model.

The number of people predicted to complete each cycle in the study is unstable given the rates of return stated in the ICR. As defined in the ICR, we have assumed the following probabilities of participating at each stage of the study. Note that the subscript $p_{x|y}$ indicates the probability of obtaining a response of x given a response in the previous Cycle of y. A value of 1 corresponds to a successful response in the Cycle while the value of 0 indicates an unsuccessful response. Responses in earlier Cycles are indicated by further symbols to the right of the solidus. For example, the symbol $p_{1|00}$ represents the probability of obtaining a successful response in a given Cycle given that the responses were unsuccessful in each of the two previous Cycles. Table A1 displays the assumed responses resulting in the original estimates for our study.

Table A1. Original estimates of response rate for study.

Participate:Previous/Next	*Yes*	*No*
Yes ($p_{1|1}$=Yes)	0.87	0.13
No last stage ($p_{1|0}$)	.16	.84
No last two stages ($p_{1|00}$)	.02	.98
No last three stages ($p_{1|000}$)	0.0	1.0

The assumptions given in Table A1 are our best estimates of the likelihood of completing the investigation. However, they are based on results from a study (Ryan et al., 1988a, Ryan et al., 1988b) for which attrition of study participants may be substantially different from that to be observed in this investigation. To evaluate the sensitivity of our results to such assumptions, we performed a sensitivity analysis varying the assumptions from those given above. Table A2 displays the result of this sensitivity analysis for the percent of participants who complete *n* cycles. The first three columns give probability of participating in *n* cycles for three different probabilities for $p_{1|1}$ given $p_{1|0}$, $p_{1|00}$, $p_{1|000}$ are the same as in Table A1. The last two columns are different probabilities for $p_{1|1}$ given that $p_{1|0}$=0.25 and other probabilities remain as in Table A1. For example, assume $p_{1|1}$= 0.8 instead of 0.87 the number of people completing all 8 cycles drops from 33 to 17. The number of individuals who complete 6 or more cycles drops from about 51 to 34. Note that assuming an improved response rate for $p_{1|0}$ increases the number of participants completing 6 or more cycle only slightly. These results call into question a simple strategy of selecting an initial population and sampling from those individuals only throughout the study.

Table A2: Probability of individual *S* participating in *n* or more cycles.

	$p_{1\|0}$=0.16			$p_{1\|0}$=0.25	
n	$p_{1\|1}$=0.87	$p_{1\|1}$=0.8	$p_{1\|1}$=0.75	$p_{1\|1}$=0.8	$p_{1\|1}$=0.8
P(*S*=0)	0.107	0.165	0.206	0.096	0.147
P(*S*≥1)	0.892	0.835	0.793	0.904	0.853
P(*S*≥2)	0.797	0.698	0.630	0.818	0.728
P(*S*≥3)	0.712	0.583	0.500	0.740	0.621
P(*S*≥4)	0.636	0.487	0.397	0.669	0.530
P(*S*≥5)	0.568	0.407	0.315	0.605	0.452
P(*S*≥6)	0.506	0.338	0.248	0.546	0.382
P(*S*≥7)	0.440	0.269	0.183	0.476	0.302
P(*S*≥8)	0.328	0.168	0.100	0.328	0.168

A second concern that arises when considering response rate is the number of participants at each cycle level. Table A3 shows the percent of individuals participating at each cycle. The top half of the table is for $p_{1|0}$=0.16 and the bottom of the table is for $p_{1|0}$=0.25 for the calculations. These calculation imply that getting people to participate over the course of the entire study is strongly correlated to the response rate, but adequate coverage may be attained if the assumptions are met.

Table A3. The fraction of individuals participating at each stage.

$p_{1\|0}$=0.16							
Cycles	2	3	4	5	6	7	8
$p_{1\|1}$=0.87	0.778	0.697	0.624	0.559	0.501	0.449	0.402
$p_{1\|1}$=0.80	0.672	0.567	0.477	0.402	0.339	0.286	0.241
$p_{1\|1}$=0.75	0.603	0.486	0.392	0.316	0.255	0.205	0.165
$p_{1\|0}$=0.25							
Cycles	2	3	4	5	6	7	8
$p_{1\|1}$=0.87	0.789	0.717	0.651	0.591	0.537	0.488	0.443
$p_{1\|1}$=0.80	0.690	0.595	0.513	0.442	0.381	0.328	0.283

One possible solution to this problem is replacing individuals who appear to drop out early in the study. Inspection of Table A1 will reveal that only a very small fraction of individuals missing two consecutive Cycles will ever return to the study. If such an individual is replaced with a new participant, it could have a substantial effect on the completed case number. Analysis shows that this could be very useful in the case that $p_{1|0}$=0.80 rather than 0.87. On the other hand, if replenishing is used and the original rate is correct it may cause a large increase in the total number of samples collected, which may put a strain on costs. Thus there is a risk trade-off to be assessed.

Table A4. Effect on total number of samples collected if number of individuals is increased to replace individuals who fail to participate in one of first two cycles.

	No replenishing pool		Replenishing pool	
probability	0.87	0.8	0.87	0.8
# samples	488	378	584	494
# successes	51	34	64	48

The probability model presented is fairly simple. This is a reasonable approach since little is known about the nature of the response rate for a study of this kind. One assumption which is likely to be incorrect is the assumption that the probabilities are stationary, i.e. they do not depend on when during the study the sequence of participation occurs. It is likely that the probability of involvement during the first cycle will be higher, while the rate of participation in the second cycle given participation in the first cycle will be lower. It is also likely that as the study nears its conclusion individuals who have participated regularly are less likely to drop out. We have performed a quick calculation using a non-stationary probability model to determine the number of individuals completing all 8 cycles given the following probabilities of participation at each of the 8 cycles (0.87, 0.5, 0.87, 0.9, 0.925, 0.95, 0.95 ,0.95). The result is 27 individuals complete 8 cycles compared to 33 for the model presented in the QSIP. This is a fairly difference given the high drop out rate assumed in the second cycle. Many such scenarios as these could be investigated. One useful result of the study will be to get estimates of what the response rate is for a study of this nature.

Precision of Parameter Estimates for Harvard/Johns Hopkins NHEXAS Investigation

Population sample size

As mentioned, population sample size is an issue when trying to estimate population parameters of exposure and the population exposure distribution. Let us focus on the population exposure distribution, e.g. the median, 90th, and 95th percentiles.

One of the main problems is one of perception. The cumulative distribution function is an sigmoidally-shaped curve and estimates of the quantiles for a specified upper percentile is inherently unstable, while estimates of the percentile of the distribution for a specified exposure level is relatively stable. If one demands high precision in the estimates for all percentiles, the required sample size is very large. On the other hand, showing that the confidence intervals for the percentiles given specified exposure level are acceptably small is much easier. This suggests that two separate evaluations of sample size are necessary. The first evaluates the precision of the quantile estimates and makes use of a simulation technique. The second evaluates the precision of the percentile estimates and uses a non-parametric approach.

Precision of Quantile Estimates Using a Parametric Approach

In order to determine whether the sample size is large enough to afford accurate and precise predictions of the "upper tail" of the distribution of exposures, we performed a series of exposure simulations as described below. There are several assumptions in this analysis. First, we assume that the sample is random. In particular, it is not subject to self-selection biases. It should be noted that any refusals can introduce biases into the sampling which are, generally, not quantifiable.

In this exercise, we have simulated results from an exposure assessment investigation with known underlying distribution. Specifically, we have assumed that the exposure (in arbitrary units) can be

characterized as a lognormal distribution with geometric mean of 10 and geometric standard deviation of 2.0. The mean value is arbitrary; the geometric standard deviation is somewhat larger than that typically found for pollutant concentrations in air, water, and soil. We investigate the stability in the estimate of several parameters in this distribution using sample sizes of 50 and 100. It should be noted that this parametric approach gives results which are clearly dependent upon the distributions selected to model the population.

The method used is from Ryan (1991) and has been used extensively in exposure assessment investigations (Ryan 1988a, 1988b, Spengler, 1994). The method invokes 30 repeated simulations of a given sample size. The row headings in the tables (Mean and StdDev) represent the mean value and standard deviation of the specific parameter of the distribution labeled as a column heading. For example, in Table A5 under the column heading 50%, the data report the result for the median or 50th percentile of the distribution. The value in the row labeled Mean (10.34) represents the mean value found for the 50th percentile averaging over the 30 simulations. The value labeled StdDev (1.34) represents the standard deviation in this value over the 30 simulations. It represents the stability of this parameter estimate. For any random sample of 50 drawn from a lognormal distribution with geometric mean of 10 and geometric standard deviation of 2, there is a 68% probability that the observed median (50th percentile) will fall within 1.34 units of the observed mean value of 10.34, i.e., $9.00 < 50\% < 11.68$.

Table A5 shows the mean value and expected variability in several parametric and non- parametric statistics for this distribution. One can note that the mean is well-predicted with an expected precision in the mean estimate of about 10%. The stability in the standard deviation, which is strongly influenced by outliers, is about 20%. A similar precision is expected in the 90th percentile estimate. This loss of precision is expected in that few data points (five) are at or above the 90th percentile suggesting a less stable estimate.

Table A5. Analysis of Variability in Predictions for a lognormal sample $\Lambda(10,2)$ with N= 50

	Mean	StdDev	25%	50%	75%	90%
Mean	12.62	9.15	6.78	10.34	16.70	25.66
StdDev	1.36	2.23	0.71	1.34	1.83	4.94

A similar analysis was performed for a sample size of 100 points. The results of this analysis are found in Table A6. Note that the mean values for each statistic are similar to those found in the 50-point simulation. The standard deviations, a measure of precision in the estimate, are smaller suggesting the expected more precise estimate of each parameter. A pattern similar to that noted in the 50-point simulation is also evident. The mean, 25th, 50th, and 75th percentiles are predicted more precisely than the standard deviation and the 90th percentile.

Table A6. Analysis of Variability in Predictions for a lognormal sample $\Lambda(10,2)$ with N= 100

	Mean	StdDev	25%	50%	75%	90%
Mean	12.80	9.58	6.57	10.45	16.16	25.12

StdDev	0.89	1.71	0.46	0.94	1.46	3.22

According to the central limit theorem, for a large number of samples, the mean value of any statistic should be normally-distributed. If this were the case in our simulations, the variance in the estimate of any statistic should be twice as large in the 50-point simulations as in the 100-point simulations. In Table A7 we present the ratio of variances found in the 50-point simulation to those found in the 100-point simulation. The values found suggest that, at least for simulations using only modestly skewed distributions, we may use this "rule of thumb" to estimate the increase or decrease in precision associated with any modification of sample size. Thus increasing the sample size form 100 to 200 would likely result in a decrease in the variance of the estimate of any parameter by a factor of two. This is equivalent to reducing the standard error or standard deviation in the estimate by a factor of the square root of two.

Table A7. Comparison of the Ratio of Variances in Selected Parameters for a lognormal sample $\Lambda(10,2)$ with sample sizes of 100 and 50.

	Mean	StdDev	25%	50%	75%	90%
Variance Ratio	2.31	1.69	2.40	2.03	1.56	2.36

To test the sensitivity of the assessment to changes in the distribution, we repeated the simulation using a more highly skewed distribution: lognormal with a geometric mean of 10 and a geometric standard deviation of 4. Tables A8-A10 present the results of the analysis. We note a deterioration of the precision overall with the Mean, 25th percentile, 50th percentile, and 75th percentile displaying 15% and 25% variability for the N=50 and N=100 cases respectively. Larger precision estimates are noted for the standard deviation and 90th percentile for reasons noted above.

Table A8. Analysis of Variability in Predictions for a lognormal sample $\Lambda(10,4)$ with N= 50

	Mean	StdDev	25%	50%	75%	90%
Mean	24.80	42.99	4.65	10.87	28.22	68.28
StdDev	7.24	26.05	0.98	2.90	6.18	27.31

Table A9. Analysis of Variability in Predictions for a lognormal sample $\Lambda(10,4)$ with N= 100

	Mean	StdDev	25%	50%	75%	90%
Mean	25.83	48.59	4.34	11.00	26.33	64.16
StdDev	5.28	21.01	0.60	1.94	4.71	16.68

Table A10 displays the variance ratios analogous to those displayed in Table A7. It is interesting to note that, even for this substantially more skewed distribution, the ratios are still near 2.

Table A10. Comparison of the Ratio of Variances in Selected Parameters for a lognormal sample $\Lambda(10,4)$ with sample sizes of 100 and 50.

	Mean	StdDev	25%	50%	75%	90%
Variance Ratio	1.88	1.54	2.69	2.22	1.72	2.68

The above discussion suggests that a sample size of 50 has the ability to give results with data quality objections sufficient to estimate population parameter.

Non-Parametric Approach to Population Sample Size

Although one may argue that it is important to shift the argument away from computing confidence intervals for the quantiles, it is also important that we can at least state the confidence intervals. Estimation using parametric distributions, such as a lognormal, will tend to highlight the problem by giving specific numbers and are highly dependent on the choice of distribution. A method for de-emphasizing the range of the confidence interval for quantiles is to use non-parametric methods.

We start by stating the problem concisely. Let p be the percentile for which the quantile is to be estimated. Then $P(X_{(r)} < q_p < X_{(s)}) = 1-\alpha$ is a $1-\alpha$ confidence interval, where $X_{(r)}$ and $X_{(s)}$ are the order statistics and $r < p < s$. Given that the observations are independent and $X_{(r)} < q_p$ if and only if at least r observations are less than q_p, the probability statement for the confidence interval has a binomial distribution. The confidence interval can be written as

$$P(X_{(r)} < q_p < X_{(s)}) = \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}$$

Table A11 gives the quantiles which are the upper and lower bounds on 90% confidence intervals for three different percentiles.

Table A11. Confidence intervals for three different percentiles.

p	size of CI	r	s
90th	0.908	42	49
95th	0.911	44	50
99th	0.910	49	

Although from a policy perspective the lower bound is more essential, it is also important to have a bounded confidence interval. Both the 90th and 95th percentiles have solid lower bounds, i.e. not near the maximum value, and bounded upper values. Only the 99th percentile confidence interval can not be bounded well in that the upper bound exceeds the maximum value.

Recall that the sampling scheme assumes that a number of people will not complete the required number of cycles to be considered a success. Unless there are arguments to be made to the contrary those individuals could also be included in the estimate of population distribution. If these individuals are included in the calculations up to 90 individuals may be used in the estimates. Such calculations

would not be that difficult and many of the issues have been addressed (Buck, 1994).

The values used in estimation of population parameters are only estimates themselves. Because of this the number of times individuals are sampled also will play a role in the quality of the estimates of population parameters. Two basic conclusions from Buck's work:

1. An adjustment factor is useful for correcting error in estimates due to influence of using estimated individual exposure and that an estimate of individual exposure variability is needed to estimate the adjustment factor.

2. Even with the adjustment factor, bias will creep into estimates of population percentile except under certain circumstances.

The arguments for choosing sample size for each individual works both ways, i.e. many good reasons for taking very few observations and many for taking a number of observations. At this point we leave the argument for choosing sample size for each individual to the next section and reiterate that the more observations which are taken the less bias introduced to the estimate of the percentiles and the more stable individual estimates of personal exposure variability.

These conclusions are based on a simple random sample with n=50. The study intends to stratify the sample into three groups of roughly equal size. The effect of this on the conclusions need to be studied. We explore this in a later section.

One further point can be made and that is the issue of estimating population distribution of these parameters. Assuming that individuals are different, the distribution of the daily variability and seasonal trends in the population is an interesting question that may be relevant to questions about acute exposure in the population. This is another reason why intra-personal estimates are important. The same arguments used for percentiles above apply and very precise estimates may not be necessary to get decent distribution estimates for these parameters.

Sample size per individual

Although it is not necessary to compute individual variances to compute a decent adjustment factor (See Buck 1994) one of the main goals of this project is to get estimates of seasonal trend and temporal variability for individuals. This goal precludes the option of using only 2 or 3 observations / individual to get estimates of population parameters. The following arguments assume that individuals have different seasonal trends and different variabilities. If these assumptions are dropped then many of the following arguments are not important and only 2 or 3 observations per individual are needed. For the discussions about sample size and testing assume three geographic groups with fifteen individuals each and five degrees of freedom (d.f.) for estimating individual variance.

Choosing the number of observations per individual is dependent on three different goals of increasing stringency.

1. Getting estimates of variability which are stable enough to give good estimates for the adjustment factor needed to estimate the population distribution.
2. Getting estimates of variability and seasonal trend with confidence intervals which are "reasonable".
3. Getting estimates of variability and seasonal trend which give reasonably powerful tests to determine whether individuals are different from one another.

The first of these goals is inherently fulfilled in most situations. If individual estimates of daily variability are used some minimal size will be required to get reasonably stable estimates of this variance which can be used in the adjustment factor. Typically, the last two goals will determine sample size.

Variance Estimation and Testing

To determine the necessary sample size for either of these goals it is useful to start by dividing the number of degrees of freedom needed to compute the seasonal trend model and the number of degrees of freedom needed to get reasonable estimates of an individuals temporal variability. If it is assumed that the seasonal trend will not be linear it is possible that 3 observations will be required to define the model. This implies that a minimum of 4 observations are required to define the model and get any estimate of temporal variability.

The next question is how many observations are needed to get a reasonable estimate of temporal variability and implicitly improve the parameter estimates of the seasonal trend model. Defining what is reasonable is fairly subjective, but easy to spot when seen. The main point is to show that the numbers will give us a useful ballpark figure which we can use as the basis for future planning. An indication of the minimum sample size needed is the confidence interval for estimates of the variance of normal random variables. Figure 1 gives the range for 95% CI on the log scale for 1 to 15 degrees of freedom for the estimate of variance. Since we will be assuming lognormal distribution these numbers are relevant if not exactly correct, the results were based on the assumption that the estimated variance is 1. If one wants to translate to another variance estimate one would need to multiply the range by the value of variance.

Figure 1 shows that at this point estimates of variance start stabilizing and the rate of return per added observation is not as large. (Although up to 12 or 13 gives a good return). We are assuming we will have 5 d.f. available for estimation of variance if 8 observations are taken per individual. The confidence interval for the variance with 5 d.f. is still fairly large ranging from $.4s^2$ to $6s^2$, but at least it is not huge.

Figure 2 shows the relative improvement in estimates of σ^2 as measured by the size of the confidence interval of the estimate. Relative efficiency is calculated as

Rel. eff.(n d.f.) = confidence interval(n-1 d.f.)/confidence interval(n d.f.).

This figure indicates that rapid improvement in estimation by increasing the number of observations by one has slowed considerably by the time the number of d.f.=5. Improvement after that is on a much more linear scale.

Figure 1. Log of range of confidence interval for estimate of σ^2 against degrees of freedom.

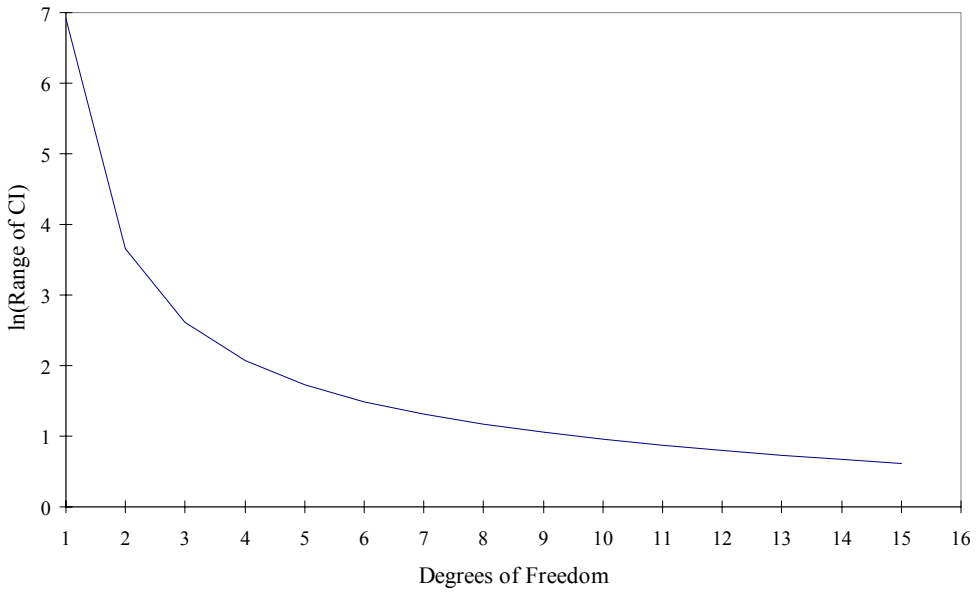
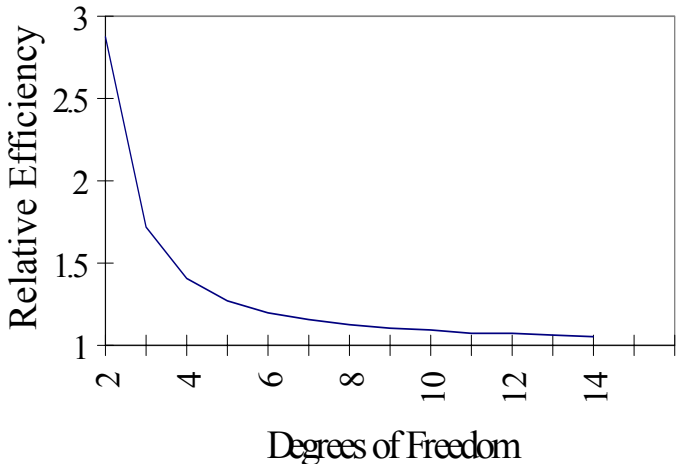


Figure 2. Improvement in estimate of σ^2 with increase in degrees of freedom of estimate.



Showing what an upper bound is for a reasonable cost efficient sample size is more difficult. Neighboring observations are likely to be correlated and hence have reduced value as an additional observation. This would reduce the number of possible observations, assuming a year long study and week long observations, to a maximum of 25. Depending on the correlation between weeks the correlation is likely to become small about three or four weeks apart which corresponds to 12 or 16 observations per individual. Showing that 8 is better than 12 or 16 is difficult. From a seasonal trend modeling point of view 8 will probably be seen as adequate, but from estimation of variance point of view it may be suspect, if seasonal trend is included.

There are two types of tests for mean and variance.

1. Do individuals have the same mean or variance.
2. Do the groups have the same mean or variance.

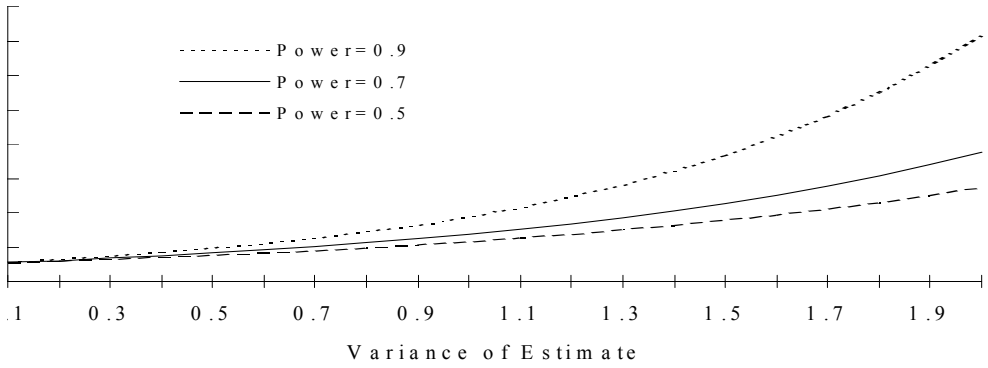
Standard ANOVA can be used to test whether the mean or variance is the same for different groups. ANOVA can be approximated for variance by using $\ln \sigma^2$ as the response variable. The rest of the analysis is pursued as it would be for analyzing the means of different groups. These tests assume that data is normally distributed, particularly the test for variance, and that the variance is the same within groups. However, the purpose here is not to carry out the actual test, but to give an indication of the power of the sample sizes chosen. These powers should roughly carry over to other tests that may be used. The power of the test for the mean and variance is given in Table A12. The power calculation is based on the assumption that two groups differ by the value Δ and all the other groups have the same mean which is the average of the two other groups, i.e. $\Delta/2$.

The power depends on the variability of the estimates of the mean and variance. This variability is defined as σ^2 in the table. For the test of the mean difference between groups the power is a linear function of this variability, e.g. the test will find absolute differences in the mean which are 2.1 times as large as σ 90% of the time. The power of the test for daily variability is a test of relative difference in daily variability. The test is in the log scale and changes non-linearly with σ. As an example Table A12 shows the relative difference detected assuming the variance in the *estimates* of daily variability = 1. Figure 3 shows the change in the power of the test for daily variability with changes in the variance in the estimate of daily variability.

Table A12: Power of test for testing between for mean and variance.

2,42 d.f.	*Δ:* In Δ/σ let $\sigma^2 = 1$;		
Δ/σ	Power	Mean	Variance
2.1	0.9	2.1	3.78
1.6	0.7	1.6	2.74
1.35	0.5	1.35	2.34

Figure 3. Change in relative difference detected in test for differences in variance between groups with change in variance in estimate of daily variability.



These calculations show that we should be able to detect reasonably small differences between groups for mean exposure. Detection of variance is more difficult to evaluate given the underlying non-linearity. One note is that these values are for relative difference in variance and it may be more relevant to report these figures as standard deviations in which case one must take square root to compute relative difference between standard deviation.

Tests for determining whether individual estimates are from the same population of estimates is equivalent to testing whether individuals have the same mean (and variance). These tests are fairly numerous, but computing the power of these tests is not as straightforward. The standard ANOVA tests can be used as they were above for testing the groups by splitting the sample for each individual into two randomly selected groups and treating them as replicates. This method is problematic in that one must decide what to do about seasonal trend. To get an indication of the power of the sample sizes chosen, this will be ignored here. Also, assume that there are six observations available for estimating variance for each individual so it can be split evenly into two groups of three observations each. Table A13 gives the results of the power calculation for this test. The problem with this test is that we have too many groups and not enough replicates of each group. This means that the chance an estimate occurs in the tail of the distribution is large and we have no resolution to distinguish it from an individual with a different mean or variance.

Table A13. Power of test to determine whether individuals have different means or variances.

44,90 d.f.	Δ: In Δ/σ let $\sigma^2 = 1$;		
Δ/σ	Power	Mean	Variance
5.75	0.9	5.75	314.2
4.65	0.7	4.65	104.6
3.83	0.5	3.83	46.1

As an example of the problem and a route to another power calculation let us look at the problem of estimating the variance. First assume that the variance is known and it is equal to one. This may be assumed because under the null hypothesis all individuals have the same variance and an estimate with 225 d.f. from the average over all individuals can be used. Although not exact, this estimate it is quite accurate. Now estimate the variance from each individual. The confidence interval for these estimates assuming σ^2=1 is (0.17,2.56). Note the difference between this confidence interval and the one for s^2. Here it is claimed that 95% of all estimates s^2 with 5 d.f. will fall in this confidence interval.

This can be considered the acceptance region for s^2. Now each individual can be considered an observation from a set of independent binomial events. Now we ask, does the estimate for the individual fall within this confidence interval? Given that we have 50 individuals, to be 90% sure that all individuals do not have the same variance more than 4 individuals have to have variance estimates outside this confidence interval. Another way to estimate the Type I error is to compute a confidence interval based on the product of 50 independent events. To guarantee a 90% confidence interval for all events the confidence interval has to be a 99.9% confidence interval for a single event. Assuming σ^2=1 the confidence interval is (0.04,4.1). If all 50 estimates fall within this range then the null hypothesis that all individuals have the same variance can not be rejected. These are fairly large numbers and help explain the problems with the ANOVA test described above.

Seasonal Trend Estimation and Testing

There are three types of test for trend that we may envision:

1. Is there a trend?
2. Do individuals have the same trend.
3. Do groups have the same trend.

Power calculation for testing the various aspects of trend which are of interest clearly depends on the

model used to estimate trend. A quick and dirty test is an ANOVA test where the cycles and groups are different treatments. This test just measures whether cycles have different means, which is enough to assume there is something going on. For our study the test would have 7 and 336 d.f. and the power of the test for is given in Table A14. These calculation show that we should be able to find changes in the mean of less than one standard deviation.

Table A14. Power calculations for two-way ANOVA test for difference in cycles.

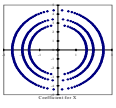
Power	Δ/σ
0.9	0.95
0.7	0.71
0.5	0.6

Power calculations for testing whether a trend exists depend on the model chosen to some extent. A quadratic model may be sufficient for unimodal seasonal trends. To start let us look at the power calculations for a single regression model $Y_i = \beta_o + \beta_1 x + \beta_2 x^2$. If we assume that the observations are evenly spaced between -1 and 1, just a matter of re-scaling the time line, then the power equation is:

$$3\phi^2\sigma^2 = 3.43\,\beta_1^2 + 1.12\,\beta_2^2$$

where φ is the power parameter determining the level of power. Figure 4 shows the ellipses for power 0.9, 0.7, and 0.5 respectively. If it is assumed that seasonal trend is fairly symmetric within a year then β_1 should be near zero. If this is the case, the figure shows that there is only a 50-50 chance of finding seasonal trends with a range of $2.75\sigma^2$ and a 90% chance of finding a seasonal trend with a range of $4.5\sigma^2$.

Figure 4. Power curves for quadratic regression equation with eight observations; curves are for power 0.9,0.7,0.5 from outside curve in respectively.



If we assume that each individual has a different regression model and want to test whether anyone has a seasonal trend the power equation is:

$$(2n+1)\phi^2\sigma^2 = 3.43\sum_i^n \beta_{1i}^2 + 1.12\sum_i^n \beta_{2i}^2$$

where *n* is the number of individuals included in the regression model. The regression model is made up of a series of columns filled with 0's except for the eight values associated with a specific individual. None of the non-zero columns overlap with non-zero columns from different individuals. This power equation does not readily explain what can be expected to be found significant and what will not since it is a 2*n* dimensional ellipse.

To look at tests whether individuals have the same trend or not, let us continue with the same model. One way to measure this is to generate the confidence interval for all separate models much as we did above for the test to see whether any trend exists. The equation for the confidence interval is also very similar:

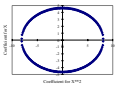
$$100\sigma^2 F(.05,100,250) = 3.43\sum_i^n (\beta_{1i} \ \hat{\beta}_{1i})^2 + 1.12\sum_i^n (\beta_{2i} \ \hat{\beta}_{2i})^2$$

assuming 50 individuals with 5 degrees of freedom for error estimates and F(.05,100,250) = 1.3. As with the test to determine whether a trend exists this is not particularly helpful in getting an idea about the size of any particular parameter.

As in the test to determine whether individuals have the same variance, if we assume it is true that all individuals have the same seasonal trend we can use all the data to get a fairly good estimate of the trend. We can then use this as our "true" value of the seasonal trend parameters and compute a 95% confidence interval for all the estimates from individuals in the sample. Figure 5 shows the range of the confidence interval. This is a conservative test and again assuming that seasonal trend is fairly

symmetric the coefficient for X^2 is the range of seasonal trend which can be expected even if they all have the same trend.

Figure 5. 95% Confidence interval for different seasonal trends assuming coefficients are known (in this case to be zero).



Stratification of Population Sample

The effect of geographic stratification on sample size for the Harvard/Johns Hopkins NHEXAS study has been questioned. Stratification has an effect on the study in two ways: through the estimates of the population parameters and through the effects on estimates and testing of intra-personal components such as daily variability and seasonal trend. Stratification actually has a positive benefit on the testing and estimation of intra-personal components because the stratification variables can be considered different levels of a factor in an ANOVA or regression model and tests and estimates of differences between the stratification levels can be made. This leaves only estimates of population parameters which may be negatively effected by using stratified sampling. The rest of this addendum looks at this issue.

For discussion purposes let us assume that the stratifications are labeled urban, suburban, and rural and that the relative size of these subpopulations is 0.45, 0.45, and 0.1 respectively. Also assume that the sample size for each subpopulation is 20, 20 and 10 respectively and the total sample size is 50 regardless of which sampling scheme is used.

To see the effect of using stratified sampling on estimates of the mean and variance of the population let us look at the equations which are used to calculate the estimates. The mean and variance are straightforward:

$$\bar{y} = \Sigma W_h \bar{y_h} \qquad \text{and} \qquad s_{st}^2 = \Sigma W_h^2 s_h^2 / n_h ,$$

where W_h is the relative size of the subpopulations and $\bar{y_h}$, s_h^2 are estimates of the mean and variance of the subpopulations.

The sample variance from a simple random sample, which is an estimate of the population variance, can be written in terms of the subpopulation variances:

$$s^2 = \sum W_h s_h^2 + \sum W_h (\overline{y_h} \overline{y})^2$$

and the variance of the estimate of the mean is:

$$s_{rs}^2 = \frac{1}{n} \sum W_h s_h^2 + \frac{1}{n} \sum W_h (\overline{y_h} \overline{y})^2 .$$

This equation shows that if all the subpopulations have the same mean and variance and if proportional sampling is used (the first term on the right is the variance of the estimate of the mean for proportional sampling) then there is no benefit to using stratified sampling. However, there is no loss either. We are not using proportional sampling, but or *ad hoc* weighting is not far off. The difference in variance between the sampling scheme we are using and proportional sampling is:

$$0.001125(s_u^2 + s_s^2) 0.001 s_r^2 .$$

The method of choice depends on the size of the variance of the subpopulations; clearly the variances need to be considerably different for the two variances estimates to be significantly different. So even if the means of the subpopulations are the same, this implies that our sampling scheme will do worse than simple random sampling only when it does worse than proportional sampling. It is expected that there are some differences between the three geographic regions so using stratified sampling is justified for estimates of the population mean.

Most of the concern about using stratified sampling is centered around getting estimates of the cumulative probability distribution function (cdf). The cdf is simply a series of binomial functions using different exposure levels as the cut-off for determining which of the two outcomes occurred. Table A15 shows the effect of using stratified sampling versus simple random sampling on the standard deviation of the estimate of different percentiles from the cdf. The table gives the true percent of the subpopulations to a given level of exposure and then the population percent exposure. The comparison is summarized in the relative efficiency of the two sampling schemes, where relative efficiency = $\sigma_{SRS} / \sigma_{st}$, the ratio of standard deviations of simple random sampling and stratified sampling. For the first five rows in Table A15, the relative efficiency is the same if the probabilities for urban and rural are switched.

Table A1. Table of relative efficiency of simple random sampling to stratified random sampling as measured by ratio of standard deviation of estimate of population percent exposure.

Urban	Suburb	Rural	Popn.	Rel. Eff.
0.5	0.5	0.5	0.5	0.970
0.4	0.5	0.6	0.465	0.978
0.3	0.5	0.7	0.43	1.003
0.2	0.5	0.8	0.395	1.053
0.1	0.5	0.9	0.36	1.142
0.85	0.9	0.95	0.8825	0.960
0.95	0.9	0.85	0.9175	0.998
0.99	0.9	0.8	0.9305	1.051
0.8	0.9	0.99	0.864	0.962
0.99	0.99	0.99	0.99	0.970

There are several noteworthy points in this table. First, stratified sampling never does much worse than simple random sampling, the worst is about 5% loss in efficiency, which seems to occur when the rural population has high exposure rates, say 95% or 99%, and the urban and suburban populations slightly lower rates, say 90%. Second, only when there are large differences between the subpopulations will stratified sampling do considerably better than simple random sampling. Third, for upper percentiles stratified sampling will do better only if the more extreme subpopulations is one of the larger subpopulations. If the rural population has the highest exposure levels then simple random sampling is better, even if there is a relatively large difference in the percent exposed for the subpopulations. If there is no difference in exposure rates between the subpopulations the relative efficiency is 0.97 and this is constant with respect to the point of the cdf being estimated. The only difficulty is that you can't use order statistics to generate the cdf.

Concerns about the use of stratified sampling are largely misplaced and there is little lost by using stratified sampling rather than simple random sampling even for the small sample sizes chosen here. The additional benefits of stratification (mainly the ability to compare these sub-populations) seem to far outweigh the small loss of accuracy that may occur by not using simple random sampling.

References

Ryan PB. Techniques in the design of large-scale field investigations. P.B. Ryan. Paper presented at the International Society for Exposure Assessment Meeting, Measuring, Understanding and Predicting Exposures in the 21st Century. Atlanta, Georgia. November 1991.

Ryan, PB, Soczek, ML, Spengler, JD, and Billick, IH. "The Boston residential NO_2 characterization study: I. A preliminary evaluation of the survey methodology." JAPCA **38**, (22-27), 1988.

Ryan, PB, Soczek, ML, Treitman, RD, Spengler, JD, and Billick, IH. "The Boston residential NO_2 characterization study: II. Survey methodology and population concentration estimates." Atmospheric Environment **22**, (2115-2125), 1988.

Spengler, JD, Schwab, M, Ryan, PB, Colome, SD, Wilson, AL, Billick, IH and Becker, E. "Personal exposure to nitrogen dioxide in the Los Angeles Basin: Study design and results." J. Air Waste Mgmt. **44**, (39-47), 1994

Buck RJ. AAAS Fellowship Final Report. September 1994.