

Strengths of the Fernald Community Cohort as a Research Resource

Medical history on this population is extremely well characterized, verified, and all cancer diagnoses have been validated with medical records. Extensive disease risk factor information has been collected, and for most information, prior to disease diagnosis.

All members of the cohort have received medical examinations, conducted at the clinical site dedicated to the program, under well-developed protocols. Conduct of examinations was monitored with various quality assurance measures. Blood pressure, pulse, height, weight, and waist circumference were taken during the examination. Physicians recorded their findings on a 10 page form designed for complete data collection, including a review of systems and a separate section for each component of the complete physical examination. The form included a section on the last 2 pages where the physician listed medical problems, and ordered program follow-up (for limited further diagnostic testing) and recorded recommendations (for further follow-up by the patient's own physician or for health education/lifestyle changes). This medical examination form was immediately reviewed by one of the nurse quality assurance specialists. Following receipt of all diagnostic testing results, a summary letter from the physician to the patient reporting all the examination findings was composed and reviewed by the examining physician. At the time of composing the letter, an extensive audit of the record was performed, to assure that all test results had been obtained, protocols for follow-up had been implemented, data recorded in the chart, and all pertinent information captured in the letter to the patient. The initial examination also included EKG's, pulmonary spirometry, and hearing testing.

Questionnaire data collection has been extensive. The initial questionnaire was a 27 page form, and yearly questionnaires averaged about 14 pages. Each year participants provided information about new medical problems and the name of the physician who had been treating them, and hospitalizations or surgeries, including the health care facility where care was received. Program nurses followed-up on this information through phone calls to participants and obtained outside medical records when indicated. Participants were also asked to list their medications, copying the name, dose and frequency from the medication container label. Participants brought all medications to their exam, so that accuracy and completeness of the medication information could be verified. All medications were assigned numerical codes on forms and in the database.



Questionnaires were designed with built in redundancy for information verification. As an example, medications and surgeries were compared with self-reported and physician-reported medical problems. Systems were in place to follow up with participants on certain health information recorded on questionnaires as well as the follow-up after medical screening. Yearly questions also asked about items such as smoking and alcohol use, diagnostic imaging, and influenza vaccinations. Each year participants also were asked a set of special questions. Examples of sets of special questions include those about family history of cancer with specific questions for each

type of relative for each cancer, physical activity, sunburns and skin moles, female reproductive history and use of oral contraceptives, and breast feeding. Complete reproductive history information has been obtained from both males and females (6 page questionnaires) and yearly pregnancy updates are obtained on all female participants of child-bearing age. At several time points during the program, participants were to complete a 2 page check list of malignant and non-malignant health conditions (and provide year of diagnosis for each condition checked). Diet history information was collected during the first and second year of the program, and extensive food frequency information (using the Harvard questionnaire) was collected during the last examination cycle. With funding from NCI, we obtained a nutrient analysis of the food frequency data.



All medical history and risk factor information is coded and stored in a series of related SAS files. All information in the chart (medical examination findings, laboratory tests, diagnostic test results, questionnaire information) is housed in a series of related SAS data files. The FMMP database consists of over 160 individual files in SAS format. All data have been entered using SAS applications designed for double data entry with verification. In addition, various computer generated

quality assurance reports are used for follow-up and to detect data outliers, which are then verified with the paper chart and the laboratory, if needed. A separate set of SAS files were created for each year of the program to maintain and manage the data collected from the physical exam forms and ongoing questionnaires. Some data collected during the program are maintained in single, cumulative SAS files rather than in separate files for each program year. Mammograms, chest x-rays, and laboratory results are examples of data maintained in single SAS files. The entire SAS database totals more than 1 gigabyte of data and spans over 10,000 variables, although many of the variables are the same from one program year to the next.

Information is coded for retrieval. All diagnoses have been assigned ICD-9 codes by a certified medical record coder. All medication information is coded by type of drug using a 4-digit code. The extensive coding of the medical information on this large cohort provides the immediate opportunity to ask specific research questions. For example, all of the 34,652 mammogram reports and the 29,759 chest x-ray reports in the database have been literally coded (coding of phrases) using a list of over 200 3-digit codes, specific to each type of imaging. Exposure or risk factor metrics have been developed from questionnaire information, and are available in the database. For example, we have developed matrix of cumulative cigarette pack-years for all participants, for each year after their year of program enrollment. We also have matrices for family history of each type of cancer for each program participant, with number of first degree and total number of blood relatives with that type of cancer.

We have collected family structure information from all participants on several different occasions and through several types of questions (family history at the time of the medial

examination, questions about who else lived with you, and questions about offspring on one of the yearly questionnaires). In each instance, we have recorded names of these relatives and coded each mentioned family member with their FMMP ID, if they are program participants. With each questionnaire we also asked participants for the names of three personal contacts, and record their relationship to the program participants. The FMMP database contains the information needed to create family pedigrees, and identify which family members are Program participants. In the near future, we hope to create family pedigrees using software such as Cyrillic or Progeny. We are familiar with the methods of pedigree creation, through Dr. Pinney's work on other genetic epidemiology studies, and have constructed a few pedigrees from this FMMP cohort. For example, through the work of a Medical Student Summer Research Program project, we easily constructed extensive pedigrees of two families with autosomal dominant polycystic kidney disease, using the information in our database and in the paper charts.

The FMMP has repeat measures on blood, serum and urine analytes on all participants, and a Medical Heritage program designed to allow standardization of the data for time dependent changes. At each examination, specimens are obtained for clinical laboratory evaluation of hematology (complete red and white blood count with differential), serum chemistry, kidney and liver function tests, FBS (fasting blood sugar), lipid profile (total cholesterol, triglyceride, LDL and HDL), TSH (thyroid stimulating hormone), and urinalysis with microscopic examination and urinary beta-2-microglobulin. Some additional testing is done on those whose test reaches a certain criterion level. For example, hemoglobin A1C is measured on all diabetics or those with a blood glucose ≥ 126 mg/dL a serum beta-2-microglobulin is measured on all those with a urine level ≥ 250 ug/L. PSA is measured in all males ≥ 50 years at the time of each examination, and yearly in those at risk of prostate cancer (25% increase over the previous PSA).

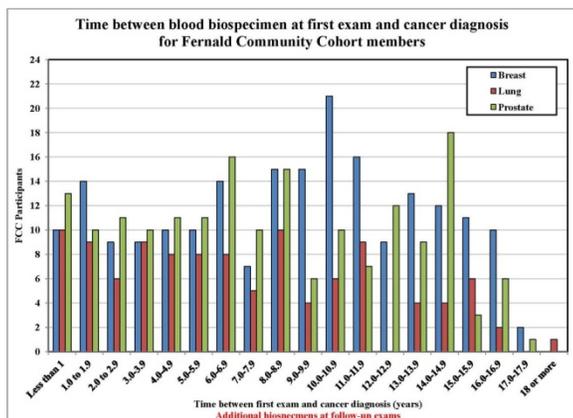


Many aspects of the program were designed to foster prospective study of the cohort. The Medical Heritage Program (designed by Drs. Amadeo Pesce and Bradley Copeland) for the stewardship of the clinical laboratory data is only one example. Not only has this program preformed quality assurance by using such measures as 30 day moving averages of the healthy Fernald participants, we have established that it is possible to track laboratory data for each analyte in a manner that will allow for adjustment for possible laboratory analytical drift found in any prospective study (Copeland and Pesce, 1992).

An extensive collection of archived samples, currently in excellent condition, already exists and is available for studies of proteomics, genetic variation, and urinary markers of both exposure and disease.

Samples of whole blood, serum, plasma and urine were obtained from all participants at the time of the initial examination and in 1997, and over 100,000 1 ml aliquots of these biospecimens have been stored at minus 80 degrees C since then. Quality of the specimens has been assessed on two occasions (primarily for degradation of protein and DNA) and found to be excellent. In 2000, the lyophilization (freeze dry) effect was

found to be 7% in two freezers, 4% in one freezer, and minimal in one freezer. In serum, albumin increased (from level measured at time of the sample collection) about 2%, and total protein about 7-11%. Since the concentrations in the archived samples were as high or higher than the initial measurements, these changes indicate concentration due to storage rather than degradation. Some enzymes showed slight levels of degradation (bilirubin, which is light sensitive; GTT; lactate dehydrogenase) but at acceptable levels. There were no decreases in AST or ALT. Of 50 whole blood samples from which DNA was isolated, only two yielded almost no DNA. For others, the DNA was of both good quantity and quality. A second more limited quality assurance analysis was conducted in 2005. Of 10 whole blood samples, the DNA isolated was of sufficient quality and quantity to be used for an array of genetic analyses (Marshall Anderson lab). More recently (2006), Dr. Shuk-mei Ho's lab found that isolated DNA was of high quality, in fragments of 15kb and greater. Dr. Detlef Shuman tested 10 serum samples for proteomic studies, comparing the samples to fresh standard human serum. None of the samples showed significant protein degradation, concentrations were in the expected range, and protein identification was consistent. Biomarkers of exposure can be measured in both blood and urine samples, and we have measured environmental biomarkers of perfluoroalkyl chemicals, bisphenol-A and urinary metals in the FMMP cohort.



Collection of many biospecimens occurred prior to disease diagnosis, and therefore the Fernald Cohort is an excellent resource for studying predictive biomarkers for disease.

The FMMP has an established procedure for applications for access to data and biospecimens. (See PDF files on this web page for the policy for acquiring data and biospecimens and the application form.) Research investigators interested in using the FCC database and samples for health-related research may apply for Access to Data and/or Biological Specimens. If approved by the Research Director, with the advice of the

Fernald Community Cohort Advisory Committee, data files from the FCC database or archived samples (frozen whole blood, serum, plasma and urine) may be distributed to qualified researchers. The University of Cincinnati and the FCC Advisory group are both very much in favor of the samples (and data) being used for research.

The extensive information on participants of the FMMP allows for the selection of tailored control groups. Criteria for controls for studies may differ depending on the disease and the lifestyle factor being studied. For example, the FMMP has provided Dr. Marshall Anderson with a set of controls for studies of identified susceptibility genetic variants related to cigarette smoking and lung cancer. For this study, the controls were persons with no first degree relatives with lung cancer, age ≥ 60 years, and with a ≥ 20 pack year

cigarette smoking history. Our large cohort and extensive database enabled us to identify more than the needed 200 controls from the group without environmental uranium exposure. The FCC also provided controls to a GWAS study conducted by Dr. Daniel Nebert, searching for “highly resistant” persons for his study of susceptibility genes for head and neck cancer, but with a ≥ 80 pack year smoking history and no history of any cancer. Persons who meet these criteria are very rare in any population, but we located 33 controls with these criteria.

The cohort is not racially diversified, which enhances the opportunity for detecting relationships between genomic variation and disease, and for discovering biomarker predictors of disease. Since 95% of the cohort is White-non-Hispanic, statistical power of genetic studies and precision of data analyses are improved by the lack of racial diversity of this cohort. Although the lack of generalizability to populations with other racial composition is certainly a limitation, the statistical associations found in this population can later be tested in other more racially diverse populations.

The database is an excellent resource for developing methods for data mining, and the potential discovery of new statistical associations will lead to hypothesis generation. The extensive database for this longitudinal cohort provides rich opportunities for data mining, and for developing and testing new methods of data mining.

Copeland BE and Pesce AJ. The medical heritage concept: a model for assuring comparable laboratory results in long-term longitudinal studies. *Annals of clinical laboratory science* 1992; 22(2):110-124.

Pinney SM, Freyberg RW, Levine GH, et al. Health effects in community residents near a uranium plant at Fernald, Ohio, USA. *International Journal of Occupational Medicine and Environmental Health*. 2003;16(2):139-153.