The New York Cancer Project: Rationale, Organization, Design, and Baseline Characteristics

Maria K. Mitchell, Peter K. Gregersen, Stephen Johnson,
Ramon Parsons, and David Vlahov for the New York
Cancer Project

ABSTRACT *Cancer is the second leading cause of death in New York City, with nearly 15,000 deaths each year. The urban setting of New York City provides ready access to large and diverse populations for whom racial/ethnic disparities in cancer risk and outcomes can be examined. A new cohort study was undertaken with several aims: (1) to provide a database and biorepository for studies of cancer etiology and pathogenesis, including host genetics; (2) to differentiate risk factors that contribute to racial/ethnic disparities in cancer risk, prevention, control, incidence, mortality, and survival; (3) to provide timely data on cancer risk and preventive behaviors that can be used to mobilize and then evaluate public health programs. Scientists from multiple institutions contributed to protocol design and implementation. Study instruments included demographics, personal and family history of cancer, risk and prevention efforts. End points include linkage with registries and medical record reviews. Using venue-based sampling with quotas, 18,187 adults aged 30 years or older were recruited over a year to undergo a baseline questionnaire, venipuncture, and contact information. The sample was 39% male, 37% older than 50 years, 58% white, 20% African American, 18% Hispanic, and 9% Asian. In terms of family history of cancer, 21% reported mother, 21% reported father, and 5.9% reported both parents with cancer; 8.5% reported any sibling with cancer. At baseline, 1,231 participants reported prior cancer. Showing the feasibility of constructing a cohort based in New York City, plans proceed for additional recruitment and analyses on the salient questions about cancer.*

KEYWORDS *Cancer, Control, Epidemiology, Etiology, Incidence, Mortality, New York City, Pathogenesis, Prevention, Survival.*

INTRODUCTION

Cancer is the second leading cause of death in the United States.[1] Although great strides have been made in treatment, underlying factors that put some individuals at higher risk for contracting the cancer are not fully characterized.[2] In the past few years, there has been a renewed focus on understanding racial and ethnic disparities in cancer incidence, mortality, and survival.[1] The New York metropolitan area is one of the most diverse and densely populated areas in the world. According to the 2000 US Census, the racial composition of the population of New York City is 44.7% white, 26.6% black or African American, 9.8% Asian, 13.5% some other race, and 4.9% more than one race.[3] In addition, 27% of the population of

Correspondence and reprints: Maria K. Mitchell, Academic Medical Development Corporation (AMDeC) Foundation, 10 Rockefeller Plaza, Suite 1120, New York, NY 10023. (E-mail: mitchell @amdec.org)

New York City consider themselves to be Hispanic or Latino. The median age of the population as of 2000 was 34.3 years.

This composition affords us a good pool from which to recruit participants for the New York Cancer Project (NYCP) to create a cohort that includes a substantial fraction of underrepresented minority groups. The American Cancer Society estimated that there were 83,700 new cases of cancer in the state of New York, with 36,200 deaths that are the result of cancer (American Cancer Society Facts and Figures, 2002, unpublished report). New York is typical of the nation in the average age-adjusted mortality rate for cancer deaths per 100,000, with New York's rate 202.3 and the national rate 206.0. New York ranks 31st highest overall in cancer mortality rates among the 50 states. By type of cancer, the highest morality is seen with lung, colorectal, breast, and prostate cancers. Differences in mortality rates by racial ethnic groups are large. For example, the average age-adjusted mortality rate for prostate cancer deaths per 100,000 men in New York from 1995 to 1999 was 29.9 for white men and 58.7 for black men. Such disparities highlight the need for a diverse cohort as a research tool to understand the underlying causes for the differences.

Cohort studies are an invaluable resource to provide information and specimens to investigate differences in incidence of cancer between populations. In the area of cancer, a number of important cohort studies have been crucial to learning about cancer. Several that have focused on cancer tended to sample nationally or statewide; these studies include the Nurses Health Study,[4] the Cancer Prevention Study II,[5] the Health Professionals Follow-up Study,[6] the Netherlands Cohort Study,[7] and the Iowa Women's Health Study.[8] Cohort studies in large urban areas are less common partly because of concerns about complexity of administration, recruitment, and retention. However, the large urban cohort permits efficient access to diverse groups (e.g., racial/ethnic, socioeconomic) who share key features of the environment, including standardized field operations.

To this end, the Academic Medical Development Company (AMDeC) Foundation was established; it is a consortium of 39 New York State institutions, including academic institutions, community hospitals, and research institutes. The full plan for the NYCP is to recruit 300,000 racially and ethnically diverse adult volunteers and to follow them for 20 years. The first phase of the project, reported here, a pilot project, had the following operational objectives: (1) recruit 17,000 racially and ethnically diverse volunteers within 24 months, (2) create a biorepository to house deoxyribonucleic acid (DNA) and plasma samples from each volunteer, and (3) create a database to store the information collected about each volunteer. The basic function of the pilot was to test the feasibility of the "urban cohort" concept and to build an infrastructure for scientists to utilize as a source of biological, behavioral, and epidemiological data.

PURPOSE

Recognizing the multidisciplinary nature of cancer studies and the need to tap into multiple resources to provide a state-of-the-art project, the NYCP has two major purposes. The first is to create an administrative structure that will attract a diverse group of scientists both to work independently on their question of interest and to consider the opportunities for collaboration and synergy with other scientists. The second purpose is to provide a core (or "skeletal") cohort structure that addresses some basic epidemiological and behavioral questions, but also is maintained as a resource for the addition of innovative scientific work using the specimens and outcome data from the cohort. For example, the technology for genetic assays has

grown considerably and likely will continue to do so; the cohort provides a framework to follow persons for cancer outcomes and then conduct nested case–control studies with the most recently developed assays relevant to pathogenesis hypotheses.

To accommodate this broad purpose, it was recognized that the design for the cohort needed to be flexible and broadly applicable for a variety of different research applications. At a minimum, the data gathered for the core cohort should address three broad areas of research activities. First, the NYCP supplies the necessary data to identify the prevalence of cancer risk behaviors and screening practices across all subgroups within the population. By following the same cohort over 20 years, the NYCP will track changes in cancer risk behaviors and screening practices over time. These behaviors can monitor the impact of such things as public health initiatives, new understanding of risk behaviors, or educational campaigns with information collected in successive waves of the annual follow-up of participants. Second, the NYCP over time will provide estimates of the incidence of and risk factors for cancer-, mortality-, and survival-based registry and vital record linkages. Third, the NYCP data will provide a repository of data, DNA material, and plasma specimens to be accessible to scientists for National Institutes of Health R01 or industry research and development applications to promote studies of etiology and pathogenesis through such mechanisms as nested case–control studies, full cohort assessment with follow-up, and requests for masked "normal controls" to permit comparison with clinical case material of the requesting investigators. A priority emphasis for the cohort is to capitalize on the ability to elucidate disparities in cancer risks.

Study Organization

The study is administered by AMDeC, but has a steering committee that includes five representatives. The steering committee includes administration from AMDeC responsible for field operations, interinstitutional relations, and budget. There is also representation from basic science, epidemiology, information systems/data management, and laboratory biorepository. This steering committee is charged with developing policies and procedures presented to a senior scientific advisory board and a community advisory board for input and expertise. The policy and procedures govern the development of a concept and protocol review committee that has a rotating list of experienced scientists for review and committees for study operations, including sampling, questionnaire development, laboratory processing and storage, quality assurance, data management and analysis (separate from the external data safety monitoring board), and external consultants for periodic ethics review of operations apart from the institutional review board.

A community advisory board was in place early to encourage rapport and responsiveness to community issues and to facilitate recruitment. As the cohort continues to form, a separate peer advisory board of volunteer study participants is part of the operations to ensure that the burden on participants or perceptions about procedures or release of results is considered with appropriate sensitivity. Meetings are regular, and ad hoc task forces are developed to address needs as they arise.

METHODS

Recruitment

Enrollment for the first phase of the NYCP occurred between January 2000 and December 2002. Fourteen enrollment sites were set up across the five boroughs of

New York City. Sites included six medical centers, two community hospitals, and six community-based health centers. In addition, the New York Blood Center enrolled individuals into the project as part of its routine donor blood drives. Enrollment sites were selected according to the demographics of the patient body in the institution and their location to ensure every borough in New York City was covered. A team of scientists (listed in the acknowledgement) developed study protocols and materials. The process was designed to recruit an ethnically and socioeconomically diverse cohort, specifically targeting individuals of African, Caribbean, Latino, Chinese, Russian, Irish, and Italian descent.

A variety of recruitment modalities were used to reach our targeted population as well as to reach high enrollment numbers. An aggressive recruitment effort included dissemination of promotional materials (developed pro bono by Young and Rubicam Advertising agency), community outreach by AMDeC staff, press conferences, advertising, and media events such as radio ads, community newspaper ads, and celebrity press events. The Young and Rubicam materials were used in two citywide advertising campaigns involving bus, subway, and convenience store displays.

Additional outreach to the community included events held with prominent political and community leaders and organizations. More locally targeted recruitment incorporated education as a main focus, including informational meetings with clinical and outreach staff at the enrollment institutions and presentations at local businesses, community centers, and community meetings. Other methods for recruitment were tailored to the needs of each enrollment site target group. Incentives were offered as a choice of two of three: a 1-day Metro card (transportation), a $10 phone card, and a T-shirt with the project logo.

Enrollment

Each NYCP enrollment site had a team of interviewers, phlebotomists, a project coordinator/outreach worker, and (for the medical centers only) a faculty member who served as the site co–primary investigator for the project. Enrollment was conducted using study-eligible volunteers either (1) on site at the enrollment centers or (2) off site in community settings where mobile site staff conduct the enrollment protocol. Interview data were collected on laptop computers and sent via a secure Internet line to the Department of Medical Informatics. Each interviewer was trained to ensure the uniform collection of data. Further, to verify adherence to interview format, AMDeC staff performed spot checks at each of the interview sites.

To be eligible, enrollees had to be 30 years of age or older, reside in the New York Tri-State area, and have a literacy level sufficient to complete a simple mail-out follow-up questionnaire (potential subjects were screened using a sample questionnaire).

Study protocol was approved by the institutional review boards of all participating institutions. All study requirements and procedures were explained to the subjects verbally and in writing, and their written informed consent was obtained. The informed consent specifically included permission to use their DNA for research purposes on a de-identified basis.

Data Collection

At baseline, consenting and eligible participants complete a screening questionnaire and a baseline questionnaire and then undergo venipuncture for 50 cc of whole blood. The questionnaire includes the following:

Demographic information: unique bar code identifier, type of encounter (screening, baseline, follow-up), site code, interviewer code, age, sex, ethnicity and race as categorized for the 2000 Census

Personal medical history: self-reported diagnoses and medical procedures for participants and their first- and second-degree relatives, with details including date and age of occurrence

Substance use: usage of tobacco, alcohol, and medications, with detail such as type, duration and frequency

Reproductive history: for women about pregnancy, menstruation, and use of hormones

Body measurements: height, weight, hip, and waist measures for the baseline visit

The instrument takes approximately one half to 1 hour to complete and was provided to interviewers on laptop computers for direct data entry. At the conclusion of each contact, the study requests (and updates) contact information such as address and telephone for subjects and alternate parties, as well as identifiers to facilitate registry linkages.

Registry linkages include the New York State department of vital records; the National Death Index; the Cancer or Tumor Registries for New York, Connecticut, and New Jersey (with further linkages to be established); and as appropriate, signed medical release forms to facilitate access to medical records. Because of the sensitivity of information being collected (including substance use), the study provides a Federal Certificate of Confidentiality issued by the Department of Health and Human Services.

Data Management

In building the information system to support the NYCP, a number of challenges had to be faced:

1. The data collection sites were geographically distributed.
2. Certain sites lacked Internet connections.
3. Interviews required considerable time to complete.
4. Participants spoke languages other than English.

Because of lack of suitable commercial solutions, the Informatics Core developed a software application that was deployed on laptops and distributed to the participating sites. Screens were designed to promote rapid data entry, easy navigation through the questionnaire, and validation logic checks in real time. Questionnaires were implemented in English, Spanish, Russian, and Chinese.

All data in the study were stored centrally. Data collected on the laptops was transmitted on a daily basis to the central server through the local network of the interview site, by phone line, and by diskette when no other means was available. A secure file transfer protocol was employed to ensure confidentiality.

Bar Coding

Another challenge for the information system was to forge a link between the subject's interview and the blood specimen for the duration of the project while protecting the subject's identity. This was accomplished through the use of bar code labels,

which were available to interviewers at all recruitment sites. Interviewers were instructed to attach identical bar code labels to the tubes used for blood samples, the consent form, and the screening form. The interviewer entered the bar code into the questionnaire software using a bar code "gun" provided with each computer.

Self-reported diagnoses and medical procedures were coded using the International Classification of Diseases (ICD). Interviewers were also permitted to enter free-form diagnoses and procedures for both subjects and family members. A program was written to parse these textual entries and assign ICD codes using standard nosology, accounting for misspellings, synonyms, and abbreviations.

Sample Management: Biorepository

The biorepository infrastructure and robotic systems were specifically designed and developed to meet the needs of the NYCP.[9] The details of the system, as well as photos and video of the laboratory operations, can be found at http://www.biorep.org. The primary focus of the biorepository is the extraction of DNA and plasma from human whole blood and the management of data associated with subsequent specimen processing.

The NYCP biorepository is designed to handle a 50-mL volume of whole blood per specimen in vacutainer tubes. Briefly, the sample processing begins with centrifugation of blood specimens with removal and storage of plasma. DNA is then purified by osmotic lysis of red blood cells with an ammonium chloride salt solution and subsequent lysis of white blood cells with a detergent lysis solution. After heating, the cell lysate protein is then precipitated using an ammonium acetate salt solution. The DNA is then precipitated with alcohol and resuspended in TRIS/EDTA buffer. Current yields from human blood are $30\,\mu g$ of DNA per milliliter of blood, and the DNA is stored at a concentration of $100\,\mu g/mL$.

A major design goal of the NYCP biorepository was to permit robotic retrieval of DNA specimens after DNA has been stored. This has been achieved through the use of a TECAN/GIRA MOLBANK storage system. The MOLBANK is a robotic freezer and specimen management system that is incorporated into the robotic operations at the North Shore Long Island Jewish Health System Biorepository. It holds 2,574 storage plates, each containing 96 DNA specimens, for a total system capacity of 247,104 specimens. The system software communicates with the MOLBANK robotic system, producing physical movement of bar-coded specimens between instruments and communication of all tracking data to a linked SQL server database. The MOLBANK system provides accurate hands-off processing of specimen storage and retrieval operations. The specimen retrieval operation of this system can be viewed at http://www.biorep.org/PublicSites/video.asp.

Statistical Analysis

Description of the baseline cohort involved frequency distributions of select variables cross-tabulated by personal history of cancer. Variables of interest for this description include those with versus without a personal history of cancer (by type), with distributions of age, sex, race/ethnicity, parental and sibling history of cancer, history of smoking, and body mass index (BMI). As a primary interest of the study involves genetic factors related to cancer, we used multivariable logistic regression to compare a family history of cancer between those with (overall) and those without a history of cancer, adjusting for age, sex, race/ethnicity, and smoking history. Odds ratios (ORs) and 95% confidence intervals (CIs) were used to guide interpretation.

RESULTS

From the recruitment, 18,187 adults older than 30 years provided complete information and specimen from venipuncture (Table 1). Of the total, 16,964 were cancer free (10,185 women and 6,771 men, and 8 with gender not recorded). The remainder reported a past history of cancer, including 96 with colorectal, 286 with skin, 88 with lymph, 96 with cervix, 57 with uterus, 409 with breast, 80 with prostate, and 187 with other cancers (combined for cancers with frequency <50 cases).

For the cancer-free group, 60.1% were female, 33.9% were at least 50 years old, 58.9% were white, 20.6% were African American, 9.6% were Asian, 7.2% were other, 3.7% refused to answer, and for ethnicity, 18.5% reported Hispanic. The distribution of racial/ethnic groups for those older than 50 years was statistically similar to the New York City 2000 Census (data not shown). In terms of family history of cancer among the cancer-free participants, 20.2% reported cancer in the mother, 20.9% in the father, 5.6% in both parents, and 7.8% in siblings. The proportion who reported a history of smoking was 41.0%. The proportion with a BMI above 25 was 65.4%, and above 30 was 27.7%.

Table 2 shows a comparison of the participants with and without a prior history of cancer by the variables in Table 1. Participants with a history of cancer were more likely than those without cancer to have a family history (in first-degree relatives) of cancer (OR 1.34; 95% CI 1.18–1.53) after accounting for age, sex, race/ethnicity, and smoking history. Because BMI may be a result rather than a predisposing factor for cancer, this was not included in the model. In separate models, the adjusted odds ratios for having a mother with cancer was 1.28 (95% CI 1.09–1.52), but was not significant for paternal history (OR 1.17) or both parents (OR 1.01). Cancer cases were more likely than cancer-free controls to have reported siblings with cancer (adjusted OR 1.56; 95% CI 1.32–1.86), and cases were also more likely to report more family members with a history of cancer (adjusted odds ratios for having one family member with cancer was 1.27; 95% CI 1.11–1.46); for two members, it was 1.38 (95% CI 1.13–1.68); and for three or more family members, it was 3.60 (95% CI 2.48–5.23).

DISCUSSION

In the past few years, the issue of racial/ethnic and socioeconomic disparities in health has become a priority national concern, and reducing such disparities is a national priority.[10] Such disparities are pronounced for large cities, where income disparities within ZIP codes can approach the bottom and the top of the scale. Andrulis[11] described the situation of health challenges in cities as the urban penalty. The NYCP provides an opportunity to acquire within a single geographic location at a single time, using the same field and laboratory methods, information between a variety of different groups with differences in cancer risk factors, screening behaviors, cancer incidence, mortality, and survival.

One of the major thrusts of the NYCP is to elucidate the genetic basis of cancer incidence. The genetic basis for cancer has been evolving (e.g., for colorectal cancer; see Ref. 12). From the questionnaire data of the baseline visit, we observed an association between prevalent cases of cancer and a family history of cancer (in first-order relatives). Although such analyses are imperfect because they involve prevalent rather than incident cases and a family history could imply a genetic source or a shared environment, these associations can be further refined to better focus laboratory investigation. Specimens from the study might also be used to investigate important environmental

TABLE 1. Baseline characteristics of New York Cancer Project by history of cancer, New York City 2000–2002.

Variable	Rectosigmoid	Skin	Lymph	Cervix	Uterus	Breast	Prostate	All Other	Cancer Free Men	Cancer Free Women	Cancer Free M+W	Total
No. of cases	93	286	88	96	57	409	80	232	6771	10185	16964	18187
sex												
% male	39.8	34.2	36.0	NA	NA	1.2	NA	35.3	NA	NA	39.9	39.0
% female	60.2	65.7	64.0	NA	NA	98.7	NA	64.5	NA	NA	60.1	61.0
Age												
% 29–39	7.6	3.9	20.2	25.2	7.2	5.4	13.9	15.7	38.0	29.6	32.9	31.4
% 40–49	21.7	25.5	30.4	35.7	29.1	18.5	8.9	24.5	33.6	32.9	33.1	32.5
% 50–59	35.8	35.8	28.1	24.2	38.1	42.3	20.2	35.1	19.9	26.1	23.7	24.5
% 60+	34.8	34.7	21.3	14.7	25.4	33.7	57.0	24.5	8.5	11.3	10.2	11.6
Race												
% white	55.4	92.9	66.2	54.2	61.1	65.6	63.2	66.3	66.1	51.8	57.5	58.2
% black	23.9	2.8	13.2	22.3	14.8	17.8	20.2	13.0	16	22.6	19.9	19.5
% Asian	9.8	1.4	4.8	1.1	7.4	6.4	5.1	11.2	6.2	12	9.5	9.2
% Other	9.8	1.4	10.8	19.1	16.6	5.1	7.6	7.1	8.8	9.8	9.4	9.2
% NA	1.1	1.4	4.8	3.1		4.9	3.8	2.2	3.0	4.1	3.6	3.6
Ethnicity												
% Hispanic	14.0	4.2	16.8	29.1	21.1	12.2	12.5	13.3	16.9	19.6	18.5	18.1
Any Cancer												
% mother	19.3	40.5	24.7	27.1	29.8	31.0	26.2	24.1	18.8	21.2	20.2	20.9
% father	22.4	33.2	27.0	24.0	36.9	27.5	13.8	22.8	21.3	20.6	20.9	21.3
% both	6.5	14.3	7.9	10.4	15.8	18.6	3.75	8.2	5.6	5.7	5.6	5.9
% any cancer (siblings)	17.2	23.1	11.2	15.6	21.1	18.6	15.0	11.6	7.5	8.5	7.8	8.5
% ever smoke	40.8	49.6	48.3	61.4	54.4	50.1	41.2	51.7	43.1	39.7	41.0	41.6
BMI												
% <18.5	1.1	2.1	0.0	2.1	1.8	1.7	0	2.2	0.3	1.2	0.9	0.9
% 18.5–<25	23.3	36.7	37.6	40	32.1	35.0	17.9	32.3	24.2	40.0	33.6	33.6
% 25–<30	45.5	36.7	45.8	32.6	37.5	37.4	61.5	35.8	48.6	30.3	37.7	37.8
% 30+	30.0	24.3	16.4	25.2	28.5	26.0	20.5	29.6	26.7	28.3	27.7	27.6

TABLE 2. Multivariate Associations for History of Cancer, New York Cancer Project

	N	OR	CL/L	CL/U	P-value
Sex					
Male	314	1.00			
Female	907	1.77	1.544	2.039	<.0001
Race					
White	821	1.00			
Black	168	0.68	0.574	0.819	<.0001
Asian	74	0.56	0.44	0.732	<.0001
Other	85	0.96	0.738	1.255	0.7770
N/A	35	1.04	0.692	1.510	0.8263
Ethnicity					
Hispanic	156	0.75	0.603	0.929	0.0212
Age group					
<=39	129	0.13	0.107	0.164	<.0001
40–49	276	0.26	0.223	0.314	<.0001
50–59	437	0.52	0.454	0.617	<.0001
60+	366	1.00			
Smoking status					
Smoker	609	1.14	1.014	1.301	0.0224
Family history of cancer					
At least one family member with cancer	664	1.34	1.188	1.527	<.0001

sources, including the etiologic role of infectious diseases and cancer. These issues may be better clarified with the prospective nature of the cohort. Specimens have been carefully selected and stored to permit focused nested case–control studies as the cohort matures and new cases of cancer are identified. A proportion of the specimens is available for investigators to test observed prevalent cases (i.e., cancer survivors) and cancer-free individuals, as well as freestanding cancer-free controls (which can be matched on a number of available variables, such as age, sex, and race/ethnicity) for investigators who have accumulated case series. To date, several such studies have been conducted.[13,14]

In addition to laboratory-based studies, this study is well positioned to document differences early in risk and screening behaviors that can call attention to deficiencies at a population level, such as the low rates of colon cancer screening.[15] The advantage of the cohort design is that follow-up questionnaires can identify rates of screening (or risk behaviors such as smoking and exercise) that can be used to evaluate communitywide campaigns to encourage higher levels of screening or to reduce disparities. Currently, cancer screening data[15,16] from the NYCP was compared with a random digit dial cross-sectional survey by the New York City Department of Health.[17] and showed highly similar associations. The behavioral dimension of the study to document change over time, although not intended as a primary aim, is an important ancillary function of the study.

Although estimates of cancer incidence and mortality require more than the numbers recruited thus far, the experience of the first phase has provided an important pilot to be able to produce a more streamlined approach in building toward the next phase of recruitment. The study is also intended to examine other outcomes, such as cardiovascular disease, diabetes, and other chronic diseases with a shorter latency period than most cancers. Thus, the study should become an important resource for a wide range of investigations critical to improving the public's health.

ACKNOWLEDGEMENT

REFERENCES

1. Arias E, Smith BL. Deaths: preliminary data for 2001. *Natl Vital Stat Rep*. 2003;51:1–44.
2. Bureau of the Census. *Census Summary Tape, File 3A (STF3A)*. Washington, DC: US Dept of Commerce; 2000
3. Govindarajan R, Shah RV, Erkman LG, Hutchins LF. Racial differences in the outcome of patients with colorectal carcinoma. *Cancer*. 2003;97:493–498.
4. Colditz GA, Manson JE, Hankinson SE. The Nurses Health Study: 20 year contribution to the understanding of health among women. *J Women's Health*. 1997;6:49–62.
5. Cancer Prevention Study II. The American Cancer Society Prospective Study. *Stat Bull Metrop Insur Co*. 1992;73:21–29.
6. Giovannucci E, Stampfer MJ, Colditz G, et al. Relationship of diet to risk of colorectal adenoma in men. *J Natl Cancer Inst*. 1992;84:91–98.
7. Van der Brandt PA, Goldbohm RA, van't Veer P, et al. A large scale prospective cohort study on diet and cancer in the Netherlands. J Clin Epidemiol 1990;43:285–95.
8. Bostick RM, Potter JD, Kushi LH, et al. Sugar, meat, and fat intake, and non-dietary risk factors for colon cancer incidence in Iowa women (United States). *Cancer Causes Control*. 1994;5:38–52.
9. Gregersen PK, Felder E. Searching for gene-environment interactions in cancer: biorepository support for the New York Cancer Project. *J Assoc Lab Automation*. 2000;5:3–5.
10. National Institutes of Health. Addressing health disparities: the NIH program of action. US. Department of Health and Human Services; 2002. Available at: http://healthdisparities.nih.gov/. Accessed November 18, 2002.
11. Andrulis DP. The urban health penalty: new dimensions and directions in inner-city health care. In *Inner City Health Care*. Philadelphia, PA: American College of Physicians; 1997.
12. Calvert PM, Frucht H. The genetics of colorectal cancer. *Ann Intern Med*. 2002;137:603–612.
13. Gruber SB, Ellis NA, Rennart G, Offit K. BLM heterozygosity and the risk of colorectal cancer. *Science*. 2002;297:2013.
14. Foulkes WD, Thiffault I, Gruber SB, et al. The founder mutation MSH2*1906G ⇒ C is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. *Am J Hum Genet*. 2002;71:1395–1412.
15. Vlahov D, Ahern J, Vazquez TL, Johnson S, Philips LA. Racial/ethnic disparities in colon cancer screening in New York City: independent of socioeconomic status, health insurance, smoking and family history. Paper presented at: Society for Epidemiologic Research meeting; June 2003; Atlanta, GA.
16. Vlahov D, Ahern J, Vazquez TL, Johnson S, Philips L, Galea S. Racial/ethnic disparities in cancer screening rates in the New York City metropolitan area. Paper presented at: American Public Health Association meeting; November 2003; San Francisco, CA.
17. New York City Department of Health and Mental Hygiene. Health alert: colonoscopy to prevent colon cancer. February/March 2003;2(2). Available at: http://www.ci.nyc.ny.us/html/doh/html/pub/pub.html. Accessed May 1, 2003.