B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

Studies of breast, thyroid, melanoma, and non-melanoma skin cancers are currently planned. Efforts to enroll additional cases reported on the Third Survey are continuing (227 breast cancer cases and 49 thyroid cancer cases are in active recruitment). About 1,200 technologists reported melanomas, 6,000 reported basal cell carcinomas, and 1,200 reported squamous cell carcinomas on the second or third follow-up surveys.

A single subset of the target population, or referent sample, will be used as a comparison group for all cancer case groups. Participation in the genetic studies component of the USRT Study requires providing a blood sample. In our previous efforts to collect blood samples from disease-free individuals for case-control studies of breast, thyroid, and other cancers (Rutter 2003 [17], Sigurdson 2005 [20], Lönn 2006 [11], Bhatti submitted [1]), approximately 48% of control subjects agreed to participate and provided a blood sample. Based on our earlier findings of enhanced participation with inclusion of a small monetary incentive (Doody 2003) [6], we will include a $2.00 bill with the blood collection kit. To accrue a referent group of 10,000 technologists, we may need to approach as many as 20,000 individuals to request their participation in the genetic studies. The universe from which participants are to be drawn is the previously-defined cohort of 110,418 technologists who were certified by the American Registry of Radiologic Technologists (ARRT) for at least 2 years between 1926 and 1980 and who

completed at least one of the two surveys conducted during the mid-1980s and mid-1990s; 35,604 technologists who did not complete either survey were excluded because radiation doses could not be computed without survey data on work history. The group being targeted for the genetic studies questionnaire and blood collection effort includes the 102,329 technologists among the 110,418 survey responders who were presumed to be alive in 1998 based on recertification with the ARRT or linkage with national mortality and other databases; excluded from the target group were: 4,170 individuals who did not work as a radiologic technologist or had an unknown date began working; 3,895 who died before 1998; 11 who were deceased with an unknown date of death (last known alive before 1998); and 13 for whom prior cancer history was unknown. We require subjects to be alive in 1998 because blood collection was begun in 1998 for selected case and control subjects. For the proposed case-cohort genetic studies, subjects will not have to be free of cancer at questionnaire response; the studies will, thus, be cross-sectional in nature. While this allows for the evaluation of prevalent cancers that occurred among the elderly (80+) early workers who potentially had the highest radiation exposures, the requirement of survival to recent years for blood collection could introduce bias. This might occur because higher levels of occupational radiation exposure are more likely before 1950 than thereafter, i.e. subjects with higher radiation exposure in the target population would tend to have survived longer than others. Since this population is unique in terms of type of radiation exposure (i.e., protracted low-dose exposures cumulating to low-to-moderate levels) and preponderance of women (73%), allowance for potential survival bias is warranted.

To maximize efficiency for assessing potential modifying effects of genetic variation on radiation-related cancer risks (i.e. gene-radiation interactions), the referent group will be over-sampled for higher ionizing radiation exposure similar to the approach employed by Cologne, *et al* (2004). Estimated doses are logarithmically distributed, with relatively few technologists having very high doses. The sampling strategy is as follows:

1. The target population of 102,329 technologists will be split into 5 categories of radiation dose (the over-sampling variable), based on Version 3.8 estimated lung doses.
2. The two highest dose categories will be 100% sampled into the referent group; this means that the 8,000 technologists with the highest doses will be included in these two categories.
3. For the three lower dose categories (cut-points to be determined based on the final lung dose distribution), 4,000 subjects from each group will be sampled into the referent group. The target population in these dose categories will first be stratified on gender (male/female) and year of birth (in 10-year categories). Within each dose category, subjects will be sampled randomly from each stratum to ensure that the distribution of all target population subjects within that dose category across strata is maintained in the sample.

Sampling weights for each dose category will be retained for use in subsequent analyses. Statistical methods and software are available to account for the different sampling fractions and appropriately adjust the variance estimates, but none currently allow for the linear excess relative risk modeling suitable for radiation exposure. We have discussed these issues with statisticians in the Biostatistics Branch, DCEG, NCI and

a likely option would be to modify the Epicure programs (Preston 1993), which were specifically designed to compute linear excess relative risks, to account for the different sampling fractions and adjust the variance estimates. Drs. Dale Preston and Jay Lubin, the co-authors of the Epicure programs, work with REB investigators on regular basis and will be available to assist in this effort.

We anticipate that power for gene-radiation interaction and breast cancer could be close to 80% for some variants, although simulations are required to more accurately estimate power given the sampling design and a continuous exposure variable. Presently we estimate the adjusted occupational radiation ERR/Gy for breast cancer in the entire cohort as 1.5 (95% CI, -0.04-4.0). For the highest category of personal diagnostic radiation exposure compared to the lowest, we have estimated an adjusted hazard ratio of 2.3 (95% CI, 1.7-3.3; assumes a 5 year lag) or 1.9 (95% CI, 1.3-2.7; assumes a 10 year lag). Using standard case-control programs for calculating power and assuming 1800 breast cancer cases and 1800 controls, a radiation odds ratio of 2.0, a gene odds ratio of 1.5, two-sided alpha of 0.05, radiation dose categories of 1, 3, 7, 12, and 20 cGy, category probabilities of 50%, 25%, 15%, 9% and 1%, at-risk genotype frequencies between from 10 to 25%, the power to detect effect modification assuming the interaction parameter on a multiplicative scale is between 1.5 to 2.5 ranges from 20 to 75%. Additional power could be gained by a control:case ratio of 2:1 rather than the 1:1 described here, using a sub-cohort referent sample stratified by dose to include a large number of high-dose comparison subjects, modeling dose continuously, assuming an additive scale for interaction, and the anticipated improvement in dose estimation. However, it is difficult to predict the amount of improvement in efficiency that might be

gained by increased precision in the dose estimates or using a dose-stratified referent sample. As mentioned earlier, we will not have sufficient power to detect gene-radiation interactions for thyroid cancer based solely on the USRT cohort; however, we hope to enroll about 1,200 papillary thyroid cancer cases in total by combining our cohort with other radiation-exposed cohorts. With 6,000 subjects reporting basal cell carcinomas in our study, we will be limited only by the costs associated with obtaining large numbers of blood samples. Our plan is to sample basal cell carcinoma cases stratified by dose in a manner similar to the referent sub-cohort selection.

2. Procedures for the Collection of Information

Optical scan data collection instruments will be used for all information gathering. Prior to mailing the survey and blood collection kit to the 20,000 technologists targeted for the referent sample, the address file will be updated using the most recent birth month renewal update information from the ARRT and address locating searches conducted by the University of Minnesota and RTI.

During Week 1, an advance letter and genetic study brochure will be mailed to subjects informing them about the study, the coming questionnaire and blood collection kit, and the planned use of the information. During Week 2, a package containing a cover letter, consent form, the genetic studies questionnaire, a blood collection kit, instructions, and a FedEx return label will be mailed to the participant. During Week 3, a follow-up letter will be mailed to all subjects. During Week 6, a follow-up phone call will be made to subjects who have not yet participated. During Week 9, a motivator postcard will be mailed to subjects who have still not responded. During Week 12, a second follow-up phone call will be made to persistent non-responders. During Week 15,

a final motivator letter will be sent to remaining non-responders. To encourage a better response rate, a $2.00 bill will be included in the initial package with the questionnaire and blood collection kit. If at any a point technologist refuses to participate, he or she will not be approached again and will be considered a non-responder in the final analysis.

Follow-up telephone calls will be made when biospecimens are received without the consent form or GQ; biospecimens will not be used until a consent form is obtained. Authorization to obtain medical validation information will be sought from subjects who previously reported physician-diagnosed cancers of interest and any who report the occurrence of such cancers during the current effort. A letter detailing the request and a HIPPA release form will be mailed to appropriate technologists with a postage-page return envelope. The medical records release forms and all procedures involved in obtaining, handling, storing, and using medical records will comply with the NCI and University of Minnesota IRB requirements and all pertinent HIPPA regulations. Follow-up mailings to non-responders will be sent approximately 3 to 4 weeks after the first mailing. Upon receiving the medical record release form, a letter plus a copy of the release form and business reply envelope will be mailed to the hospital or physician of record. Follow-up mailings and telephone calls will be made to those who don't respond. For subjects who live in selected states with accredited cancer registries, cancer validation will be sought directly from the registries since they will have already confirmed the cancers as primary malignant tumors and performed nosology coding according to the International Classification of Diseases for Oncology.

3. **Methods to Maximize Response Rates and Deal With Non-response**

Cohort members selected for the subcohort who are currently lost-

to-follow-up will be traced through a variety of tracing resources to obtain current contact information. These resources may include, but not be limited to: the Social Security Administration databases; TeleMatch, a computerized residential telephone number look-up service accessing over 65 million listings, including over one million not-yet-published numbers of new movers; LexisNexis, which provides batch matching against data from all public records including Driver's License, Department of Motor Vehicle, and voter registration; National Change of Address (NCOA), a service provided by the US Postal Service, which can be searched by batch to identify new addresses for recent movers with service updates of almost 100 million records every 2 weeks and storage of new address information for 18 months; and proprietary credit bureau databases such as Equifax. Technologists who cannot be found will be classified as non-responders.

As described in 2. above, one week before questionnaire mailings, advance letters will be sent with a genetic studies brochure that describes the study in detail, answers frequently asked questions, and provides contact information for additional questions or to opt out. One week after questionnaire mailings, a follow-up letter will be sent, followed by a follow-up telephone call at 6 weeks, a motivator postcard at 9 weeks, a second follow-up telephone call at 12 weeks, and a final motivator letter at 15 weeks, as needed.

To improve communications and provide greater reciprocity with cohort members, detailed study newsletters were sent in 2001 and 2004, 2-page study updates were sent in 2005 and 2006 to all study participants detailing the research objectives and progress to date, and we plan to send a newsletter in 2007. A study-specific internet homepage (http://dceg2.cancer.gov/radtechs/) provides subjects ready access to research

summaries, published papers, study newsletters, responses to frequency asked questions, and other study information. .

4. Test of Procedures or Methods to be Undertaken

Based on initial interviews, minor modifications were made to a few questions to improve clarity and no further problems were encountered.

The completed interview forms were reviewed and scanned for data capture. Quality assurance reviews were conducted to verify the accuracy of the scanning, to identify any problem areas to be addressed during interviewer training, and to finalize the edit/coding procedures. A 100% validation of the scanned data against the original forms was completed and final revisions were made to the Genetic Studies Questionnaire (Attachment 2). The scanning programs were updated for the revised form and another 100% validation of the scanning process was performed for a sample of forms. A Decision Log table was created to document all edit/coding decisions. The same version of the questionnaire will be sent by mail and we do not anticipate that subjects will have difficulty completing the mail survey.

To test likely response by control subjects to receipt of a mailed questionnaire and blood collection kit, we are currently conducting a pilot test of 800 high-dose controls. This test will be completed by the end of February 2007. If response to the mail solicitation is inadequate (less than ~50%), we will reconsider the use of telephone solicitation either in lieu of, or following, a failed mail solicitation. Budgetary constraints provide the impetus behind our plan to conduct the solicitation by mail, if at all possible.

5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or

Analyzing Data

The University of Minnesota is under contract to the NCI (NO1-CP-31018) to collaborate in this research effort. The University of Minnesota Principal Investigator is Bruce H. Alexander, PhD, Division of Environmental and Occupational Health (telephone 612-625-7934, email balex@umn.edu). RTI International is under contract to the NCI (N02-CP-31013) to provide support services, e.g. cohort follow-up, for this and other Radiation Epidemiology Branch investigations. The RTI Project Director is John P. Heinrich, MPH (telephone 301-230-4642, email jheinrich@rti.org). Information Management Services, Inc. (IMS) is under contract to the NCI (N02-CP-31003) to provide statistical and analytic support for this research effort. The IMS Project Director is Janice Beach (telephone 301-680-9770, email beachj@imsweb.com). The Project Officers from the Radiation Epidemiology Branch, NCI who are responsible for overseeing the data collection and analyzing the data are Michele M. Doody, MS (telephone 301-594-7203, email doodym@exchange.nih.gov) and Alice J. Sigurdson, PhD (telephone 301-594-7911, email sigurdsa@mail.nih.gov). Statisticians within the Epidemiology and Biostatistics Program, Division of Cancer Epidemiology and Genetics, NCI who have provided statistical consultation and collaboration on this study are Drs. Barry Graubard (301-496-7455), Deukwoo Kwon (301-451-4348) and Charles Land (301-594-7165). They and others (e.g. Drs. Dale Preston (707-476-8648) and Jay Lubin (301-496-3356)) will be consulted throughout the study as necessary.