Background Statement

The Radiation Epidemiology Branch (REB) of the Division of Cancer Epidemiology and Genetics (DCEG) requires the support for all phases of studies that include the coordination of multi-center studies in the U.S. and abroad, facilitation of cooperation between NCI and its collaborators; development of efficient data management systems; the design of data collection documents; hiring, training and supervision of interviewers and abstractors; collecting, keying, editing, updating and re-coding data; tracing individuals for vital status; obtaining radiation treatment records; validating cancer incidence and mortality; collecting biologic and environmental specimens and shipping them to appropriate facilities for storage and analysis; analyzing or arranging for analysis of these specimens; and creating and manipulating data files.

REB conducts a cutting-edge, high-impact research program designed to identify, understand, and quantify the risk of cancer in populations exposed to different types of radiation. The research program includes novel epidemiologic studies conducted in the U.S. and abroad that integrate radiation exposure, risk factor information, and at times biologic markers to provide new insights into cancer etiology and mechanisms of action. Key research areas within REB's broad research program include: 1) Cancer risks associated with radiation exposure from medical sources, both diagnostic and therapeutic; 2) Cancer risks associated with lower dose radiation exposure from occupational and environmental sources; and 3) Etiology of highly radiosensitive malignancies (cancers of the thyroid, brain, breast, hematopoietic malignancies and others). Radiation dosimetry is a key component of radiation risk assessment, and REB dosimetrists are closely integrated within epidemiologic studies to provide dose estimates and associated uncertainties, assist in study design, and collaborate on analysis of almost all REB epidemiologic investigations.

The support services contractor will provide support (data collection, quality control, and obtaining biologic and environmental specimens) to multiple Branch studies. The contract will serve as a mechanism that will allow REB to address rapidly emerging issues of public health or clinical concern associated with radiation exposure from medical, occupational and environmental sources and cancer risk. In the past, REB has been called upon to launch studies of public health concern (e.g., cellular telephones and risk of brain tumors, electromagnetic fields and risks of childhood leukemia), and the Branch needs to be able to quickly mobilize resources. The contractor shall make appropriate personnel available as requirements for support change over time. The contract will support at least 8 long-term ongoing projects that will span several years (e.g., risks and determinants of radiation-related second cancers after breast cancer in HMO settings, cancer risks among persons in Belarus exposed *in utero* to fallout from the Chernobyl accident, second primary cancers after radiotherapy for retinoblastoma, and dosimetry methodologic studies). In addition, the contractor will support pilot studies to investigate the feasibility of undertaking novel investigations of cancer related to radiation exposure that incorporate distinctive genetic and dosimetry components (e.g. second cancer risks after intensity modulated radiotherapy and protons, pilot study for a US biobank cohort using HMOs that will be a Division-wide resource). It is estimated that at least 4 newly initiated studies will need support within the first two years of the new contract.

Appendix A lists the studies that are currently supported by the incumbent support services contractor, the type of support activity for each study, and the size of the effort. The table also indicates which studies will be continued under the new contract that will require support. A bibliography of the published papers from epidemiologic studies that were supported by the current contractor is also attached (Appendix B).

The current 5-year support services contractor is RTI International, and they have been the incumbent since 2009 (HHSN261200900017C).

Appendix A.

This appendix includes ongoing and proposed new studies to be supported within the first two years of the new support services contractor, grouped by REB key research areas. The size of the effort by the Contractor is indicated at the end of the descriptive paragraph as:

Small	**=<$100,000**
Medium	**=≥$100,000-$500,000**
Large	**=≥$500,000**

These costs represent the estimated costs needed for support services for each study from contract award to completion of the study or the end of the contract period, whichever comes first.

<u>Ongoing Studies:</u>

Medical Radiation Exposure and Second Cancer Risk

1. **Second Cancers After Breast Cancer in Kaiser HMO**
   Breast cancer survivors now make up about 25% of the 11 million cancer survivors living in the U.S. Evaluation of the role of radiation, chemotherapy, and other factors that could impact their long-term health is of clinical and public health importance. Historically, most studies of these late effects have been case-control studies that have required extensive efforts to reconstruct treatment histories from paper medical records. NCI has been evaluating newer, more cost-effective approaches, and recently completed a pilot study of breast cancer survivors in the Kaiser Northwest HMO. The pilot demonstrated that it was feasible to generate a historical cohort of 4717 breast cancer survivors, diagnosed from 1991 to 2009, with comprehensive exposure information by linking multiple electronic health records that included rich information on treatments (including all hormonal therapy prescriptions, chemotherapies and radiotherapy), smoking histories, height, weight as well as extensive details on the cancer diagnoses and other outcomes. In the past year, NCI added a second cohort of 3646 breast cancer survivors from Kaiser Colorado to create a large cohort of breast cancer survivors (n=8363) with an average of 10 years post-diagnosis follow-up that will enable evaluation of long-term second cancer risks, in particular lung cancer, in relation to a variety of breast cancer therapies and lifestyle variables including obesity and smoking. **It is anticipated that the Contractor shall be the successor in interest for the subcontract with Kaiser HMOs. The Contractor shall coordinate subcontracts with these Kaiser HMOs in order to continue to follow the cohort for the incidence of second cancers survivors and to coordinate the calculation of radiation doses to the lung that the cohort received as part of their breast cancer treatment, as well as provide general oversight and quality control of data received from Kaiser. (Medium)**

   There are no SOPs available for this study.

2. **Second Cancer Risk Following Chemotherapy and Radiotherapy for Retinoblastoma**
   Since 1984, REB investigators have been evaluating subsequent cancer risk in a unique cohort

of 2,000 1-year retinoblastoma (Rb) survivors. NCI identified increased risks of second cancers in the cohort that have been attributed to the interaction of radiotherapy and genetic susceptibility. Given the current median age of the Rb cohort (38 yrs), the very high risk for second cancers in patients with hereditary Rb, and the presence of the *RB1* mutations in the pathway of several common epithelial adult tumors, and the newer treatments for Rb, NCI plans to continue to follow the cohort over the next decade to quantify cancer risks. The cohort of survivors was interviewed by telephone in 1993, 1998, 2001, and 2009. The most recent National Death Index search was conducted in 2011. Approximately 35% of the cohort has developed a second cancer. All of the second cancers have been confirmed by pathology reports whenever possible. **The Contractor shall continue to trace the cohort for vital status and current address and telephone number in case NCI decides to conduct another telephone survey of the cohort. The Contractor shall continue to collect information on second cancers and conduct National Death Index searches, confirm and code second cancers, update the study cohort files and prepare analytic data files. (Medium)**

The following SOPs are available: (located after Appendix B)

- Medical Record Abstraction Form
- Training Manual for Abstractors
- Questionnaire
- Cancer Confirmation Process

Occupational Radiation Exposure

3. **U. S. Radiologic Technologist (USRT) Cohort studies: Dose Reconstruction: Nuclear Medicine Technologists and Technologists Performing/ Assisting with Fluoroscopically-Guided Procedures**
   The USRT study began in 1984 by NCI and is a nationwide retrospective cohort of 146,020 radiologic technologists who were certified between 1926 and 1984. Four surveys have been conducted in this cohort to ascertain work history, medical history, and various cancer risk factors, such as lifetime UV exposure. A major emphasis during the last few years has been a retrospective dose reconstruction to assign estimated annual and cumulative occupational radiation doses to individual technologists to assess occupational radiation-related dose-response patterns. However, absorbed doses to the technologists who assist with nuclear medicine (NM) procedures may be higher than those of other radiologic technologists, because radionuclides have much higher energies with greater tissue penetration than lower energy x- or γ-rays. Similarly, doses may be substantially higher for technologists who assist with fluoroscopically-guided interventional (FG) procedures due to the type and length of the procedures compared with radiography procedures. NCI is currently collecting additional work history information to improve dose estimation for workers who performed NM and FG procedures. REB dosimetrists will develop detailed estimation of annual and cumulative dose estimates for radiologic technologists in the cohort whose occupational history indicates primarily working in NM or assisting or performing FG procedures. **It is anticipated that the Contractor shall be the successor in interest for the subcontract with USRT. The Contractor shall support expert dosimetry and statistical consultants and update vital status and film badge dose data through linkage with the American Association of**

Radiologic Technologists and a commercial dosimetry company. (Medium)

There are no SOPs available for this study. The Contractor shall provide mathematical, scientific and computer consulting/support for epidemiologic research for REB. The Contractor shall work directly with REB staff and others on various aspects of the analysis and management of data from the x-ray technologists study. A particular emphasis on this work will be the development and practical implementation of statistical methods to incorporate information on dosimetric uncertainty.

Environmental Radiation Exposure

4. **Dose Reconstruction and Health Effects of *In-utero* and Early-life Radiation Exposure (Chernobyl-related studies)**
NCI is currently collaborating with investigators in Belarus to establish a cohort of about 2,500 individuals exposed *in utero* and in early life to radioactive iodine and other radioactive isotopes from Chernobyl fallout, reconstructing doses to the thyroid gland from internal and external radiation exposure, ascertaining thyroid cancer and other thyroid diseases and assessing radiation dose-responses for these endpoints. NCI is working with the collaborators to locate cohort subjects, interview their mothers, collect personal information on places and dates of residence, consumption of local foodstuff, countermeasures applied (e.g., administration of stable iodine), time spent outdoors, construction materials of houses, etc., as well as specifics regarding the timing of the pregnancy during the period of exposure. NCI will use this information for dose reconstruction. Measurements of 131-I content in the mother's thyroid gland have been made and they will be linked to subjects to reconstruct fetal thyroid dose. Comprehensive dose reconstruction will take into account both internal and external sources of irradiation. The collaborators will ascertain thyroid cancer by linking the cohort with the Belarus National Cancer Registry. **The Contractor shall be the liaison between the local public health officials in Belarus who are collaborating on this project. The Contractor shall coordinate the various phases of this project to assist with data management, pre-test questionnaires as needed, and ensure that milestones are met in a timely fashion and within budget and ship tissue specimens (about 100) to a NCI designated laboratory for analysis. (Medium)**

There are no SOPs available for this study. The Contractor shall provide support on computer codes for calculating thyroid doses due to iodine-131 intakes using a published dosimetry model. The Contractor shall adapt information obtained from Iodine-131 and Cesium-137 deposition and questionnaire data and run computer code to calculate instrumental and ecological doses to the thyroid for cohort members. The Contractor shall obtain reliable central estimates and stochastic thyroid doses due to exposure to Iodine-131 for cohort members. To evaluate the feasibility of data collection for military clean-up workers, the Contractor shall collect data with detailed descriptions of 'clean-up' activities for military clean-up workers including results of direct thyroid measurements that were taken shortly after the Chernobyl accident in 1986.

Dosimetry Studies

5. **Medical Radiation Dosimetry for Epidemiologic Studies**
   In order to quantify the risk of cancer in populations exposed to ionizing radiation, REB investigators frequently require accurate estimates of organ-specific radiation dose in large numbers of individual study subjects in order to estimate dose-response relationships. Dosimetry support will include making measurements on anthropomorphic/water/other phantoms, abstracting information on radiotherapy in medical records, calculating radiation doses to organs of interest as well as uncertainties in these dose estimates, and providing advice in support of pilot studies on the quality of radiotherapy data and probable estimates or organ doses for specific radiation exposures. **An example of future REB-led project that shall require dosimetry support is Kaiser Permanente Health Plan Study to examine the long-term cancer risks of breast cancer treatments in a cohort of 8000 women. It is anticipated that the Contractor shall be the successor in interest for the subcontract with Kaiser. Another example is an ongoing study of gene-radiation interactions on second cancer risk in the Childhood Cancer Survivor Cohort (CCSS). The Contractor shall provide dosimetry support for approximately 500 survivors for CCSS. The Contractor shall work closely with extramural medical physicists and other dosimetry staff to provide support for a program of dose estimation for REB's ongoing and new epidemiologic studies. (Large)**

   There are no SOPs available for this study. The Contractor provides services to support REB research activities in the area of radiation dosimetry in support of health risk studies.

6. **Support state-of-the-art computer program for individual organ dose calculations from radiological accidents or terrorist events**
   REB is currently developing a computer program for external organ dose estimations for individuals that might be exposed to photon radiation from radiological accidents, emergencies, or terrorist activities. The software program will include three components: (1) a Monte Carlo radiation transport code especially designed for human anatomy with relatively simple radiation field geometries, (2) realistic human anatomical models ranging from newborn to adult males and females that are flexible enough to represent any posture and body size of interest during exposure, and (3) a graphical user interface (GUI) for the user to specify the type and geometry of the radiation source and to specify simple shielding geometries that might have protected the person during exposure, and to conveniently calculate absorbed doses to individual organs. The software will calculate absorbed doses to organs through radiation transport simulations. **The Contractor shall coordinate with consultants to test the software as it proceeds through stages including testing support of the Monte Carlo code, the GUI, and extensive comparison to other dose estimation tools as well as support of experimental validation. The Contractor shall also ensure that testing milestones are met on time and within budget. (Small-Medium)**

   There are no SOPs available for this study.

Multi-center Studies

7. **InterLymph Subtypes International Project**
   Non-Hodgkin lymphoma (NHL), which affects over 86,000 individuals in the United States annually, is composed of numerous closely-related yet heterogeneous diseases with distinctive morphologic, immunophenotypic, genetic, and clinical features. Because individual epidemiologic studies do not have sufficient sample size to evaluate risk factors by NHL subtype, the International Lymphoma Epidemiology Consortium (InterLymph) has undertaken a consortium-wide initiative, the "InterLymph NHL Subtypes Project" with the primary aims of 1) comprehensively evaluating risk factor profiles across NHL subtypes in a multivariate setting, and 2) quantitatively assessing etiologic heterogeneity across NHL subtypes. The project pools individual-level data on over 15 lifestyle, medical history, family history, and occupational risk factors from 20 case-control studies across North America, Europe, and Australia, including a total of 17,564 NHL cases and 24,109 controls. Data are being analyzed centrally and will be published as 13 manuscripts in a special issue of a journal. **The Contractor shall provide logistical support by organizing email communications amongst over 50 collaborators on 3 continents, including scheduling and coordinating approximately 2 teleconferences per year as needed, preparing and distributing teleconference notes to all participants, and tracking key study tasks. (Small)**

   The following SOPs are available: (located after Appendix B)
   - List of collaborators
   - Monograph proposal

8. **Periconception Folic Acid Supplements and Other Postulated Risk Factors for Pediatric Leukemia and Other Childhood Cancers: A Record-Linkage Study in a Chinese Cohort**
   The Chinese Center for Disease Control and Prevention (CCDC), the National Cancer Institute (NCI), and U.S. Centers for Disease Control and Prevention (CDC) investigators will evaluate prospectively the role of periconception folic acid supplements and other postulated risk factors in risk of developing pediatric leukemia and other cancers. This study, the Children and Families Cohort Study (CFCS), will be carried out in a unique population of offspring of Chinese mothers who were enrolled during 1993-95 in a large (N=243,779) Community Intervention Program (CIP) of periconception folic acid supplements. **The Contractor shall coordinate the various phases of this project to oversee data collection, assist with data management, and ensure that milestones are met in a timely fashion and within budget. A pilot study is now underway to determine if a main study is feasible. The aims for the pilot study include identifying and abstracting medical records for 1) all pediatric cancer cases in children born during 1994-96 and diagnosed in the 8 pilot study hospitals during 1994 to the present, (2) all pediatric cancer cases or related disorders that are incompletely worked up for diagnosis of pediatric cancer and/or completely worked up but not treated among children born during 1994-96. The aims of the Main Study are: (1) to compare risk of pediatric leukemia, brain tumors, and other cancers diagnosed in offspring of mothers who took versus did not take folic acid supplements during the peri-conception/early first trimester period based on the 1993-95 CIP records and (2) to assess prospectively postulated prenatal, early neonatal,**

and pre-conception risk factors for pediatric leukemias, brain tumors, and all cancers combined in the CFCS cohort using interview data collected from mothers and medical records during 1993-96 in a hospital and community-based component. (Medium)

The following SOP is available: (located after Appendix B)

- Chinese language data manual

NEW STUDIES

1. **Second Cancer Risks after Intensity Modulated Radiation Therapy and Proton Therapy**
   Intensity modulated radiation therapy (IMRT) is a newer and rapidly expanding approach to the planning and delivery of radiotherapy. IMRT allows dose to be concentrated in the tumor volume and so should result in better tumor control and reduction of acute toxic effects in surrounding organs. However, because more monitor units and more fields are used than in conventional radiotherapy the result is a larger total-body radiation dose and dose to organs outside the radiation fields. Hence, the reduction in acute toxic effects may be at the cost of increased risks of second malignancies. To date there have been no direct studies of the long-term second cancer risks from IMRT use. Proton therapy is also rapidly expanding and many similar concerns exist for this alternative therapy, especially exposure to neutrons as a by-product. NCI plans to conduct feasibility work to explore strategies for constructing a cohort of patients treated with these conventional and newer therapies in the U.S and to estimate doses. **The Contractor shall conduct feasibility work including evaluation of treatment data completeness at state cancer registries and extraction of samples of medical and radiotherapy records. For these efforts the Contractor shall develop coding and data extraction forms. (Large)**

2. **Cancer in Solid Organ or Bone Marrow Transplant Patients – Expansion with More Recently Treated Patients**
   Patients undergoing bone marrow transplantation (BMT) are of particular interest due to their unusual combination of exposures, including total body irradiation with 8-15 Gy, intensive chemotherapy, and late complications of severe immune dysfunction. To address the late carcinogenic effects of high dose radiotherapy in the presence of severe immune dysfunction, NCI assembled an international cohort of nearly 29,000 patients who received allogeneic BMT between 1964 and 1996 at 272 transplant centers worldwide. NCI proposes to (1) conduct a feasibility study to assemble a new cohort that will evaluate new transplantation procedures and (2) increase follow-up for new malignancies for the existing cohort through 2004. There are several novel areas of study that NCI proposes to add during the expansion of this cohort, including evaluation of the relation between viral infections (e.g., cytomegalovirus) and subsequent malignancies, and a study to assess the feasibility of collecting biospecimens for future studies of genetic susceptibility to malignancies following transplantation. **The Contractor shall conduct feasibility work and coordinate a follow-up of this cohort. (Medium)**

3. **A Pilot Study for a US Biobank Cohort Using HMOs**
   Many of the existing etiology cohorts in NCI are now 15-20 years past their initial baseline,

have limited remaining biologic samples and are aging. There is the need, therefore, to start investigating settings for a new generation of cohorts that could answer a broad range of etiology questions and include multiple biologic samples for longitudinal molecular studies. The HMOs provide a rich potential setting for such a cohort because of the passive follow-up systems via high-quality cancer registries, electronic medical records with data on medications, radiotherapy, co-morbidities and intermediate end-points and centralized laboratories for processing samples. In the next few years, NCI plans to conduct feasibility and then pilot studies for a U.S. biobank cohort within a group of HMOs. This is envisioned to be a Division-wide resource. **The Contractor shall help identify collaborating centers, assess the best collection method for biological samples, and conduct pilot work for questionnaire development. If the pilot work is successful, then the Contractor may provide support with similar components of the full study. (Small)**

4. **U.S. Veterans "Weatherman" Biodosimetry Study**
REB will investigate the use of quantitative biodosimetry (biomarkers of radiation exposure, including fluorescent in situ hybridization) for subjects exposed many decades before the assay, as well as summarize the cancer and mortality experience of U.S. veterans with the highest recorded radiation exposures. The proposed study will: (1) Validate the use of biodosimetry for exposures that took place six decades in the past. The exposures are estimated to range from 300 mGy-1 Gy; (2) Compare the biologically-based dose estimates with those from film badges; and (3) Summarize other data for two highly exposed groups of veterans. The subjects of this study will be (i) 9 military air weathermen (9 of 27 are still alive) exposed on Rongerik Atoll in the Pacific in 1954 to fallout from the largest U.S. nuclear test, and (ii) the 9 living military personnel who were observers to nuclear tests in 1952 at the Nevada Test Site and who received the highest doses according to the Dept. of Defense (DoD). **The Contractor shall manage all aspects of the project. They shall provide support for development of the questionnaire, pre-test the questionnaire, contact subjects and collect questionnaire data and biosamples, and ship samples to a NCI designated laboratory. The Contractor shall prepare an analytic data file. (Small)**

Appendix B. Bibliography

PUBLICATIONS SINCE THE LAST Senior Advisory Group (SAG) REVIEW (Period of current contract July 2009- June 2014)

1A. Medical, Diagnostic Radiation

CT Scan Use

1. **Berrington de González A**, Kim KP, Smith-Bindman R, McAreavey D. Myocardial perfusion scans: projected population cancer risks from current levels of use in the United States. Circulation. 2010 Dec 7; 122(23):2403-10.
2. **Berrington de González A**, Kim KP, Yee J. CT colonography: perforation rates and potential radiation risks. Gastrointest Endosc Clin N Am. 2010 Apr; 20(2):279-91.
3. **Berrington de González A**, Kim KP, Knudsen AB, Lansdorp-Vogelaar I, Rutter CM, Smith-Bindman R, Yee J, Kuntz KM, van Ballegooijen M, Zauber AG, Berg CD. Radiation-related cancer risks from CT colonography screening: a risk-benefit analysis. AJR Am J Roentgenol. 2011 Apr; 196(4):816-23.
4. **Berrington de González A**, Mahesh M, Kim KP, Bhargavan M, Lewis R, Mettler F, **Land C**. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Arch Intern Med. 2009 Dec 14; 169(22):2071-7.
5. **Berrington de González A**. Estimates of the potential risk of radiation-related cancer from screening in the UK. J Med Screen. 2011; 18(4):163-4.
6. Kim KP, Einstein AJ, **Berrington de González A**. Coronary artery calcification screening: estimated radiation dose and cancer risk. Arch Intern Med. 2009 Jul 13; 169(13):1188-94.
7. **Linet MS**, Slovis TL, Miller DL, Kleinerman R, Lee C, **Rajaraman P**, **Berrington de González A**. Cancer risks associated with external radiation from diagnostic imaging procedures. CA Cancer J Clin. 2012; 62:75-100.

Conference on Radiation and Health

8. Stram D, Puskin, J, Redpath, L, Kronenberg A, Brooks T, Schubauer-Berigan M, Preston D, **Kleinerman RA**, **Sigurdson A**, Zablotska L, Ronckers C, Colman M, **Little M**. New developments and future directions in radiation research American Statistical Association Conference on radiation and health. Vail, Colorado, June15-18, 2008. Radiation Research. 2010; 173, 392-398.
9. Stram D, **Sigurdson A**, Schubauer-Berigan M, Darby S, Zimbrick J, Colman M, Bhatti P, Cullings H, Kronenberg A, Amundson S, Jacob P, Puskin J. Radiation exposure in medicine: biological and public health significance American Statistical Association Conference on radiation and health. Annapolis, Maryland, June 13-16, 2010. Radiation Research. 2011; 175, 131-142.

Scoliosis Study

10. Ronckers CM, **Land CE**, Miller JS, Stovall M, Lonstein JE, **Doody MM**. Cancer mortality among women frequently exposed to radiographic exams for spinal disorders. Radiat Res. 2010; 174(1):83-90.

1B. Medical, Therapeutic Radiation

Benign Gynecologic Disorders

11. Sakata R, **Mabuchi K**, **Kleinerman RA**, Stovall M, Smith SA, Thompson WD, Wactawski-Wende J, Cookfair DL, Hadjimichael O, Boice JD Jr, **Inskip PD**. Cancer mortality following radiotherapy for benign gynecological disorders. Radiat Res, 2012, 178;266-79.

Retinoblastoma Survivors

12. Bhagia P, Colanta A, Abramson DH, Carlson D, **Kleinerman RA**, Kraus D, Dunkel I. Sinonasal adenocarcinoma: A rare second malignancy in long-term retinoblastoma survivors. Pediatr Cancer Blood. 2011; 57:693-695.
13. Francis JH, **Kleinerman RA**, Seddon JM, Abramson DH. Increased risk of secondary uterine leiomyosarcoma in hereditary retinoblastoma. Gynecol Oncol. 2012; 124:254-9.
14. **Kleinerman RA**, **Schonfeld SJ**, Tucker MA. Sarcomas in hereditary retinoblastoma. Clinical Sarcoma Review. 2012. Epub ahead of print.
15. **Kleinerman RA**, Yu CL, **Little MP**, Li Y, Abramson DH, Seddon JH, Tucker MA. Variation of second cancer risk by family history among long-term survivors of retinoblastoma. J Clin Oncol. 2012;30:950-957
16. **Kleinerman RA**, Yu CL, **Little MP**, Li Y, Abramson DH, Seddon JH, Tucker MA. Reply to AC Moll, et al. J Clin Oncol. 2012;24:3028-29.
17. **Kleinerman RA**. Radiation sensitive genetically susceptible pediatric sub-populations. Pediatr Radiol. 2009; 39 (Suppl 1):S27-S31.
18. Liu J, Givi B, Wolden S, **Kleinerman RA**, Dunkel IJ, Lee N, Shah JP, Abramson DH, Kraus DH. Secondary skull base malignancies in survivors of retinoblastoma: The Memorial Sloan Kettering Cancer Center experience. Skull Base Journal. 2011; 21:103-107.
19. Mills MB, Balise RR, Hudgins L, **Kleinerman RA**. Mutation risk associated with paternal and maternal age in a cohort of retinoblastoma survivors. Hum Mutat. 2012; 131:1115-1122.
20. Yu CL, Tucker MA, Abramson DH, Seddon JM, Stovall M, Fraumeni JF Jr, **Kleinerman RA**. Cause-specific mortality in long-term survivors of retinoblastoma. J Natl Cancer Inst. 2009:101:581-91.

Second Gastrointestinal Cancers

21. Chaturvedi AK, **Kleinerman RA**, Hildesheim A, **Gilbert EA**, Chen BE, Storm H, Lynch CF, Hall P, Langmark F, Pukkala E, Kaijser M, Andersson M, Fossa S, Joensuu H, Travis LB, Engels EA. Second cancers following squamous cell carcinoma and adenocarcinoma of the cervix. J Clin Oncol, 2009; 27:967-73.
22. **Kleinerman RA**, Smith SA, Holowaty E, Hall P, Pukkala E, Vaalavirta L, Stovall M, Weathers R, **Gilbert E**, Aleman B, Kaijser J, Andersson M, Storm H, Joensuu H, Lynch CF, Dores GM, Travis LB, **Morton LM**, **Curtis RE**. Radiation dose and subsequent risk for stomach cancer in long-term survivors of cervical cancer. IJROBP. 2012. Submitted.
23. **Morton LM**, **Gilbert E**, Hall P, Andersson M, Joensuu H, Dores GM, Stovall M, Holowaty EJ, Lynch CF, **Curtis RE**, Smith SA, **Kleinerman RA**, Storm H, Pukkala E, Weathers R, **Linet MS**, **Rajaraman P**, Fraumeni JF, Brown LM, van Leeuwen FE, Fossa SD, Johannesen TB, Langmark F, Travis LB, Aleman B. Risk of treatment-related esophageal cancer among 5-year breast cancer survivors. Ann Oncol. Epub on line, June 27, 2012.

2. Environmental and occupational exposure

Benzene-Exposed Workers in China

24. Lan Q, Zhang L, Shen M, Jo WJ, Vermeulen R, Li G, Vulpe C, Lim S, Re X, Rappaport SM, Berndt SI, Yeager M, Yuenger J, Hayes RB, **Linet M**, Yin S, Chanock S, Smith MT, Rothman N. Large-scale evaluation of candidate genes identifies associations between DNA repair and genomic maintenance and development of benzene hematotoxicity. Carcinogenesis. 2009;30:50-8.
25. Zhang L, Lan Q, Ji Z, Li G, Shen M, Vermeulen R, Guo W, Hubbard AE, McHale CM, Rappaport SM, Hayes RB, **Linet MS**, Yin S, Smith MT, Rothman N. Leukemia-related chromosomal loss detected in progenitor cells of benzene-exposed workers. Leukemia 2012 May 30; doi:10.1038/supportsuppleu.2012.143.

Chernobyl Research Studies

26. Abend M, Pfeiffer RM, Ruf C, **Hatch M**, Bogdanova TI, Tronko MD, Riecke A, Hartmann J, Meineke V, Boukheris H, **Sigurdson A**, **Mabuchi K**, **Brenner A**. Iodine-131 dose dependent gene expression in thyroid cancers and corresponding normal tissue following the Chernobyl accident. PlosOne. 2012. In press.
27. Bozhok Y, Greenebaum E, Bogdanova TI, McConnell RJ, Zelinskaya A, **Brenner AV**, Zurnadzhy LY, Zablotska L, Tronko MD, **Hatch M**. NA cohort study of thyroid cancer and other thyroid diseases after the Chernobyl accident: cytohistopathologic correlation and accuracy of fine-needle aspiration biopsy in nodules detected during the first screening in Ukraine (1998-2000). Cancer Cytopathology. 2009;117(2):73-81.
28. **Brenner AV**, Tronko MD, **Hatch M**, Bogdanova TI, Oliynik VA, Lubin JH, Zablotska LB, Tereschenko VP, McConnell RJ, Zamotaeva GA, O'Kane P,

Bouville AC, Chaykovskaya LV, Greenebaum E, Paster IP, Shpak VM, **Ron E**. I-131 Dose-response for incident thyroid cancers in Ukraine related to the Chornobyl accident. Environ Health Perspect. 2011;119:933-9.

29. **Cahoon EK**, Rozhko A, Hatch M, Polyanskaya O, **Ostroumova E**, Tang M, Nadirov PE, Yauseyenka V, Savasteeva I, McConnell RJ, Pfeiffer RM, **Brenner AV**. Factors associated with serum thyroglobulin levels in a population living in Belarus. Clinical Endocrinology. In press.
30. Cardis E, **Hatch M**. The Chernobyl accident--an epidemiological perspective. Clin Oncol (R Coll Radiol) 2011 May;23(4):251-60.
31. **Drozdovitch V**, Minenko V, Khrouch V, Leshcheva S, Gavrilin Yu, Khrutchinsky A, Kukhta T, Kutsen S, Luckyanov N, Shinkarev S, Tretyakevich S, Trofimik S, Voillequé P, **Bouville A**. Thyroid dose estimates for a cohort of Belarusian children exposed to radiation form the Chernobyl accident. Radiat Res. In press.
32. **Drozdovitch V**, Zhukova O, Germenchuk M, Khrutchinsky A, Kukhta T, Luckyanov N, Minenko V, Podgaiskaya M, Savkin M, Vakulovsky S, Voillequé P, **Bouville A**. Database of meteorological and radiation measurements made in Belarus during the first three months following the Chernobyl accident. J Environ Radioact 116, 84 – 92; 2013.
33. **Hatch M**, **Brenner A**, Bogdanova T, Derevyanko A, Kuptsova N, Likhtarev I, **Bouville A**, Tereshchenko V, Kovgan L, Shpak V, **Ostroumova E**, Greenebaum E, Zablotska L, **Ron E**, Tronko M. A screening study of thyroid cancer and other thyroid diseases among individuals exposed in utero to iodine-131 from Chernobyl fallout. J Clin Endocrinol Metab 2009;94(3):899-906.
34. **Hatch M**, Furukawa K, **Brenner AV**, Olinjyk V, **Ron E**, Zablotska LB, Terekhova GM, McConnell RJ, Markov VV, Shpak VM, **Ostroumova E**, **Bouville AC**, Tronko M. Prevalence of hyperthyroidism after exposure during childhood or adolescence to radioiodines from the Chornobyl nuclear accident: dose-response results from the Ukrainian-American Cohort study. Radiat Res. 2010 Dec;174(6):763-72.
35. **Hatch M**, Polyanskaya O, McConnell R, Gong Z, **Drozdovitch V**, Rozhko A, Prokopovich A, Petrenko S, **Brenner A**, Zablotska L. Urinary Iodine and Goiter Prevalence in Belarus: experience of the Belarus-American cohort study of thyroid cancer and other thyroid diseases following the Chernobyl nuclear accident. Thyroid. 2011 Apr;21(4):429-37.
36. Khrutchinsky A, **Drozdovitch V**, Kutsen S, Minenko V, Khrouch V, Luckyanov N, Voilleque P, **Bouville A**. Mathematical modeling of a survey-meter used to measure radioactivity in human thyroids: Monte Carlo calculations of device response and uncertainties. Appl Radiat Isotopes. 2012;70: 743–751.
37. Khrutchinsky A, Kutsen S, Minenko V, Zhukova O, Luckyanov N, **Bouville A**, **Drozdovitch V**. Monte Carlo modeling of beta-radiometer device used to measure milk contaminated as a result of the Chernobyl accident. Appl Radiat Isot. 2009; 67:1089-93.
38. Kryuchkov V, Chumak V, Maceika E, Anspaugh LR, Cardis E, Bakhanova E, Golovanov I, **Drozdovitch V**, Luckyanov N, Kesminiene A, Voillequé P, **Bouville A**. RADRUE method for reconstruction of external doses to Chernobyl liquidators in epidemiological studies. Health Physics. 2009; 97(4):275-298.

39. Leeman-Neill RJ, **Brenner AV**, **Little MP**, Bogdanova TI, **Hatch M**, Zurnadzy LY, **Mabuchi K**, Tronko MD, Nikiforov YE. RET/PTC and PAX8/PPAREγ chromosomal rearrangements in post-Chernobyl thyroid cancer and their association with I-131 radiation dose and other characteristics. Cancer. 2012. In Press.
40. Likhtarev I, Kovgan L, Masiuk S, Talerko M, Chepurny M, Ivanova O, Gerasymenko V, Boyko Z, Voilleque P, **Drozdovitch V**, **Bouville A**. Thyroid cancer study among Ukrainian children exposed to radiation after the Chornobyl accident: improved estimates of the thyroid doses to the cohort members. Health Physics. Submitted.
41. Likhtarov I, Kovgan L, Chepurny M, Ivanova O, Boyko Z, Ratia G, Masiuk S, Gerasymenko V, **Drozdovitch V**, Berkovski V, **Hatch M**, **Brenner A**, Luckyanov N, Voilleque P, **Bouville A**. Estimation of the thyroid doses for Ukrainian children exposed in utero after the Chernobyl accident. Health Physics 2011;100(6):583-593.
42. **Mabuchi K**, **Hatch M**, **Little MP**, **Linet MS**, **Simon SL**. Risk of thyroid cancer after adult radiation exposure: Time to reassess? Radiat Res. 2012. In press.
43. Nadyrov E, Rozhko A, Kravtsov V, **Mabuchi K**, **Hatch M**, Nakamura N, Nikonovich S, Aleksanin S. Karyopathological traits of thyrocytes and exposure to radioiodines in Belarusian children and adolescents following the accident at the Chernobyl nuclear power plant. Radiat Environ Biophys 2012;51: 187-93.
44. **Neta G**, **Hatch M**, **Kitahara CM**, Tronko M, **Brenner A**. In utero exposure to iodine-131 from Chernobyl fallout and growth in adolescence. Submitted.
45. O'Kane P, Shelkovoy E, McConnell RJ, Shpak V, Parker L, **Brenner A**, Zablotska L, Tronko M, **Hatch M**. Frequency of undetected thyroid nodules in a large I-131-exposed population repeatedly screened by ultrasonography: results from the Ukrainian-American cohort study of thyroid cancer and other thyroid diseases following the Chernobyl accident. Thyroid. 2010 Sep;20(9):959-64.
46. **Ostroumova E**, **Brenner A**, Oliynik V, McConnell R, Robbins J, Terekhova G, Zablotska L, Likhtarev I, **Bouville A**, Shpak V, Markov V, Masnyk I, **Ron E**, Tronko M, **Hatch M**. Subclinical hypothyroidism after radioiodine exposure: Ukrainian-American cohort study of thyroid cancer and other thyroid diseases after the Chornobyl accident (1998-2000). Environ Health Perspect 2009;117(5):745-50.
47. **Ostroumova E**, Rozhko A, **Hatch M**, Furukawa K, Polyanskaya), McConnell R, Nadyrov E, Petrenko S, Romanov G, Yauseyenka V, **Drozdovitch V**, Minenko V, Prokopovitch A, Savasteeva I, Zablotska L, **Mabuchi K**, **Brenner A**. Functional thyroid diseases among children and adolescents in Belarus exposed to Iodine-131 from the accident at the Chernobyl nuclear plant. Environ Health Persp. 2012. In Press.
48. Skryabin AM, **Drozdovitch V**, Belsky Y, Leshcheva SV, Mirkhaidarov AK, Voillequé P, Luckyanov N, **Bouville A**. Thyroid mass in children and adolescents living in the most exposed areas to Chernobyl fallout in Belarus. Radiat Prot Dosim. 2010;142: (2-4): 292-299.
49. Tronko M, **Mabuchi K**, Bogdanova T, **Hatch M**, Likhtarev I, **Bouville A**, Oliynik V, McConnell R, Shpak V, Zablotska L, Tereshchenko V, **Brenner A**, Zamotayeva G. Thyroid cancer in Ukraine after the Chernobyl accident (in the framework of Ukraine-U.S. Thyroid Project). J Radiol Prot 32:N65-N69; 2012.

50. Zablotska L, **Ron E**, Rozhko AV, **Hatch M**, Polyanskaya ON, **Brenner AV**, Lubin JH, Romanov GN, McConnell RJ, O'Kane P, Evseenko VV, **Drozdovitch V**V, Luckyanov N, Minenko VF, **Bouville A**, Masyakin V. Thyroid cancer risk in Belarus among children and adolescents exposed to radioiodine after the Chornobyl accident. British Journal of Cancer 2011:104, 181-187.
51. Zablotska LB, Bazyka D, Lubin JH, Gudzenko N, **Little MP**, **Hatch M**, Finch S, Dyagil I, Reiss RF, Chumak VV, **Bouville A**, **Drozdovitch V**, Kryuchkov VP, Golovanov I, Bakhanova E, Babkina N, Lubarets T, Bebeshko V, Romanenko A, **Mabuchi K**. Radiation and the risk of chronic lymphocytic and other leukemias among Chornobyl cleanup workers. Environmental Health Perspectives. 2012. In press.
52. Zvonova I, Krajewski P, Berkovsky V, Ammann M, Duffa C, Filistovic V, Homma T, Kanyar B, Nedveckaite T, **Simon SL**, Vlasov O, Webbe-Wood D. Validation of models for 131I ecological transfer and thyroid radiation dose using Chernobyl fallout data. Journal of Environmental Radioactivity. 2010; 101(1):8-15.

Interventional Fluoroscopists

53. Kim KP, Miller DL, **Berrington de González A**, Balter S, **Kleinerman RA**, **Ostroumova E**, **Simon SL**, **Linet MS**. Occupational radiation doses to operators performing fluoroscopically guided procedures. Health Physics. 2012; 103:80-99.

NCI Conference: Late Health Effects of Ionizing Radiation

54. Jacob P, **Ron E**. Late health effects of ionizing radiation: bridging the experimental and epidemiological divide. Radiat Environ Biophys. 2010; 49:109–110.
55. Krestinina L, Preston DL, Davis FG, Epifanova S, **Ostroumova E**, **Ron E**, Akleyev A. Leukemia incidence among people exposed to chronic radiation from the contaminated Techa River, 1953–2005. Radiat Environ Biophys. 2010; 49:195–201.
56. **Little MP**, Tawn EJ, Tzoulaki I, Wakeford R, Hildebrandt G, Paris, F, Tapio S, Elliott P. Review and meta-analysis of epidemiological associations between low/moderate doses of ionizing radiation and circulatory disease risks, and their possible mechanisms. Radiat Environ Biophys. 2010; 49:139–153.
57. **Ron E**, Jacob P. Late health effects of ionizing radiation: bridging the experimental and epidemiologic divide. Radiation Research. 2012; 174, 789–792.
58. **Linet MS**, Kim KP, Miller DL, **Kleinerman RA**, **Simon SL**, **Berrington de González A**. Historical Review of Occupational Exposures and Cancer Risks in Medical Radiation Workers. Radiation Research. 2012; 174(6b):793-808.
59. Preston DL, Yu Krestinina L, Sokolnikov ME, **Ron E**, Davis FG, Ostroumova, EV, **Gilbert ES**. How Much Can We Say about Site-Specific Cancer Radiation Risks? Radiation Research. 2012;174(6b):816-824.
60. **Ron E**, **Brenner A**. Non-malignant thyroid diseases after a wide range of radiation exposures. Radiation Research. 2012; 174(6b):877-888.

U.S. Radiologic Technologists

61. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, Eccles D, Evans DG, Fletcher O, Johnson N, dos Santos Silva I, Peto J, Stratton MR, Rahman N, Jacobs K, Prentice R, Anderson GL, Rajkovic A, Curb JD, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver WR, Bojesen S, Nordestgaard BG, Flyger H, Dork T, Schurmann P, Hillemanns P, Karstens JH, Bogdanova NV, Antonenkova NN, Zalutsky IV, Bermisheva M, Fedorova S, Khusnutdinova E, Kang D, Yoo KY, Noh DY, Ahn SH, Devilee P, van Asperen CJ, Tollenaar RA, Seynaeve C, Garcia-Closas M, Lissowska J, Brinton L, Peplonska B, Nevanlinna H, Heikkinen T, Aittomaki K, Blomqvist C, Hopper JL, Southey MC, Smith L, Spurdle AB, Schmidt MK, Broeks A, van Hien RR, Cornelissen S, Milne RL, Ribas G, González -Neira A, Benitez J, Schmutzler RK, Burwinkel B, Bartram CR, Meindl A, Brauch H, Justenhoven C, Hamann U, Chang-Claude J, Hein R, Wang-Gohrke S, Lindblom A, Margolin S, Mannermaa A, Kosma VM, Kataja V, Olson JE, Wang X, Fredericksen Z, Giles GG, Severi G, Baglietto L, English DR, Hankinson SE, Cox DG, Kraft P, Vatten LJ, Hveem K, Kumle M, **Sigurdson A**, **Doody M**, Bhatti P, Alexander BH, Hooning MJ, van den Ouweland AM, Oldenburg RA, Schutte M, Hall P, Czene K, Liu J, Li Y, Cox A, Elliott G, Brock I, Reed MW, Shen CY, Yu JC, Hsu GC, Chen ST, Anton-Culver H, Ziogas A, Andrulis IL, Knight JA, Beesley J, Goode EL, Couch F, Chenevix-Trench G, Hoover RN, Ponder BA, Hunter DJ, Pharoah PD, Dunning AM, Chanock SJ, Easton DF. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet. 2009; 41(5):585-90.
62. Aschebrook-Kilfoy B, **Neta G**, **Brenner AV**, Hutchinson A, Pfeiffer RM, Sturgis EM, Xu L, Wheeler W, **Doody MM**, Chanock SJ, **Sigurdson AJ**. Common genetic variants in metabolism and detoxification pathways and the risk of papillary thyroid cancer. Endocr Relat Cancer. 2012; 19(3):333-44.
63. **Berrington de González A**, Hartge P, Cerhan JR, Flint AJ, Hannan L, MacInnis R, Moore SC, Tobias GS, Anton-Culver H, Freeman LB, Beeson WL, Clipp SL, English DR, Folsom AR, **Freedman DM**, Giles G, Hakansson N, Henderson KD, Hoffman-Bolton J, Hoppin JA, Koenig KL, Lee I-M, **Linet MS**, Park Y, Pocobelli G, Schatzkin A, Sesso, HD, Weiderpass E, Willsox BJ, Wolk A, Zeleniuch-Jacquotte A, Willet WC, Thun MJ. Body mass index and mortality among 1.46 million white adults. N Engl J Med. 2010; 363(23):2211-9.
64. Bhatti P, **Doody MM**, **Rajaraman P**, Alexander BH, Yeager M, Hutchinson A, Thomas G, Hunter DJ, **Simon SL**, Weinstock RM, Rosenstein M, Stovall M, Preston DL, **Linet MS**, Hoover RN, Chanock SJ, **Sigurdson AJ**. Novel breast cancer risk alleles and interaction with ionizing radiation among U.S. radiologic technologists. Radiat Res. 2010; 173(2):214-24.
65. Bhatti P, Kampa D, Alexander BH, McClure C, Ringer D, **Doody MM**, **Sigurdson AJ**. Blood spots as an alternative to whole blood collection and the effect of a small monetary incentive to increase participation in genetic association studies. BMC Med Res Methodol. 2009 Nov 13; 9:76.

66. Bhatti P, Yong LC, **Doody M**, Preston DL, Kampa DM, Ramsey MJ, Ward EM, Edwards AA, **Ron E**, Tucker JD, **Sigurdson A**. Diagnostic X-ray examinations and increased chromosome translocations: evidence from three studies. Radiat Environ Biophys. 2010; 49(4):685-92.
67. **Brenner A**, **Neta G**, Hutchinson A, **Ron E**, Pfeiffer R, Aschebrook-Kilfoy BA, Abend M, Sturgis E, Li G, Wheeler W, Chanock S, **Sigurdson AJ**. Common genetic variants in immunity and the risk of thyroid cancer. PLoS One. (in press).
68. **Cahoon E**, **Rajaraman P**, Alexander BH, **Doody MM**, **Linet MS**, **Freedman DM**. Use of non-steroidal anti-inflammatory drugs and risk of basal cell carcinoma in the United States Radiologic Technologists Study. Int J Cancer. 2012; 130(12):2939-48.
69. Daugherty SE, Pfeiffer RM, **Sigurdson AJ**, Hayes RB, Leitzmann M, Schatzkin A, Hollenbeck AR, Silverman DT. Nonsteroidal anti-inflammatory drugs and bladder cancer: A pooled analysis. Am J Epidemiol. 2011; 173(7):721-30.
70. Dunning AM, Healey CS, Baynes C, Maia AT, Scollen S, Vega A, Rodriguez R, Barbosa-Morais NL, Ponder BA, Low YL, Bingham S, Haiman CA, Le Marchand L, Broeks A, Schmidt MK, Hopper J, Southey M, Beckmann MW, Fasching PA, Peto J, Johnson N, Bojesen SE, Nordestgaard B, Milne RL, Benitez J, Hamann U, Ko Y, Schmutzler RK, Burwinkel B, Schurmann P, Dork T, Heikkinen T, Nevanlinna H, Lindblom A, Margolin S, Mannermaa A, Kosma VM, Chen X, Spurdle A, Change-Claude J, Flesch-Janys D, Couch FJ, Olson JE, Severi G, Baglietto L, Borresen-Dale AL, Kristensen V, Hunter DJ, Hankinson SE, Devilee P, Vreeswijk M, Lissowska J, Brinton L, Liu J, Hall P, Kang D, Yoo KY, Shen CY, Yu JC, Anton-Culver H, Ziogoas A, **Sigurdson A**, Struewing J, Easton DF, Garcia-Closas M, Humphreys MK, Morrison J, Pharoah PD, Pooley KA, Chenevix-Trench G. Association of ESR1 gene tagging SNPs with breast cancer risk. Hum Mol Genet. 2009; 18(6):1131-9.
71. Fasching PA, Pharoah PD, Cox A, and 106 others including **Rajaraman P, Schonfeld SJ.** The role of breast cancer susceptibility variants as prognostic factors. Hum Mol Genet, 2012, 21:3826-39.
72. **Freedman DM**, **Cahoon EK**, **Rajaraman P**, Major J, **Doody MM**, Alexander BH, Hoffbeck RW, Kimlin M, Graubard BI, **Linet MS**. Sunlight and other determinants of circulating 25-hydroxyvitamin D levels in black and white participants of a nationwide U.S. study. Am J Epidemiol. In press.
73. **Freedman DM**, Kimlin MG, Hoffbeck RW, Alexander BH, **Linet MS**. Multiple indicators of ambient and personal ultraviolet radiation exposure and risk of non-Hodgkin lymphoma (United States). J Photochem Photobiol B. 2010;101(3):321-5.
74. Freedman DM, **Rajaraman P**, Fuhrman B, Hoffbeck R, Alexander BH. Sunlight, hormone replacement status and colorectal cancer risk in post-menopausal women. Int J Cancer 2010;126(8):1997-2001.
75. Fuhrman BJ, **Freedman DM**, Bhatti P, **Doody MM**, Fu Y-P, Chang S-C, **Linet MS**, **Sigurdson AJ**. Polymorphisms in vitamin D-related genes and risk of breast cancer. Submitted.
76. Gaudet MM, Milne RL, Cox A, Camp NJ, Goode EL, Humphreys MK, Dunning AM, Morrison J, Giles GG, Severi G, Baglietto L, English DR, Couch FJ, Olson JE, Wang X, Chang-Claude J, Flesch-Janys D, Abbas S, Salazar R, Mannermaa A,

Kataja V, Kosma VM, Lindblom A, Margolin S, Heikkinen T, Kampjarvi K, Aaltonen K, Nevanlinna H, Bogdanova N, Coinac I, Schurmann P, Dork T, Bartram CR, Schmutzler RK, Tchatchou S, Burwinkel B, Brauch H, Torres D, Hamann U, Justenhoven C, Ribas G, Arias JI, Benitez J, Bojesen SE, Nordestgaard BG, Flyger HL, Peto J, Fletcher O, Johnson N, Dos Santos Silva I, Fasching PA, Beckmann MW, Strick R, Ekici AB, Broeks A, Schmidt MK, van Leeuwen FE, Van't Veer LJ, Southey MC, Hopper JL, Apicella C, Haiman CA, Henderson BE, Le Marchand L, Kolonel LN, Kristensen V, Grenaker Alnaes G, Hunter DJ, Kraft P, Cox DG, Hankinson SE, Seynaeve C, Vreeswijk MP, Tollenaar RA, Devilee P, Chanock S, Lissowska J, Brinton L, Peplonska B, Czene K, Hall P, Li Y, Liu J, Balasubramanian S, Rafii S, Reed MW, Pooley KA, Conroy D, Baynes C, Kang D, Yoo KY, Noh DY, Ahn SH, Shen CY, Wang HC, Yu JC, Wu PE, Anton-Culver H, Ziogoas A, Egan K, Newcomb P, Titus-Ernstoff L, Trentham Dietz A, **Sigurdson AJ**, Alexander BH, Bhatti P, Allen-Brady K, Cannon-Albright LA, Wong J, Chenevix-Trench G, Spurdle AB, Beesley J, Pharoah PD, Easton DF, Garcia-Closas M. Five polymorphisms and breast cancer risk: results from the Breast Cancer Association Consortium. Cancer Epidemiol Biomarkers Prev. 2009; 18(5):1610-6.

77. Gerstenblith MR, **Rajaraman P**, Khaykin E, **Doody MM**, Alexander BH, **Linet MS**, **Freedman DM**. Basal cell carcinoma and anthropometric factors in the U.S. Radiologic Technologists Cohort Study. Int J Cancer. 2012; 131(2):E149-55.
78. Howard RA, Leitzmann MF, **Linet MS**, **Freedman DM**. Physical activity and breast cancer risk among pre- and postmenopausal women in the U.S. Radiologic Technologists cohort. Cancer Causes Control. 2009; 20(3):323-33.
79. Jiao L, **Berrington de González A**, Hartge P, Pfeiffer RM, Park Y, **Freedman DM**, Gail MH, Alavanja MC, Albanes D, Beane Freeman LE, Chow WH, Huang WY, Hayes RB, Hoppin JA, Ji BT, Leitzmann MF, **Linet MS**, **Meinhold CL**, Schairer C, Schatzkin A, Virtamo J, Weinstein SJ, Zheng W, Stolzenberg-Solomon RZ. Body mass index, effect modifiers, and risk of pancreatic cancer: a pooled study of seven prospective cohorts. Cancer Causes Control. 2010; 21(8):1305-14.
80. **Khaykin Cahoon E**, Wheeler DC, Kimlin MG, Kwok R, Alexander BH, **Little MP**, **Linet MS**, **Freedman DM**. 2012. Individual, environmental, and climatic predictors of daily personal ultraviolet radiation exposure measurements in a United States study. PLoS ONE. In press.
81. **Kitahara CM**, **Linet MS**, Beane Freeman LE, Check DP, Church TR, Park Y, Purdue MP, Schairer C, **Berrington de González A**. Cigarette smoking, alcohol intake, and thyroid cancer risk: A pooled analysis of five prospective cohorts in the United States. Cancer Causes Control. 2012; 23(10):1615-24.
82. **Kitahara CM**, Platz EA, Freeman LE, Hsing AW, **Linet MS**, Park Y, Schairer C, Schatzkin A, Shikany JM, **Berrington de González A**. Obesity and thyroid cancer risk among U.S. men and women: a pooled analysis of five prospective studies. Cancer Epidemiol Biomarkers Prev. 2011; 20(3):464-72.
83. Kwok RK, **Linet MS**, Chodick G, **Kleinerman RA**, **Freedman DM**, Fears T, Johnson RE, Alexander BH. Simplified categorization of outdoor activities for male and female U.S. indoor workers-a feasibility study to improve assessment of ultraviolet radiation exposures in epidemiologic study questionnaires. Photochem Photobiol. 2009; 85(1):45-9.

84. Lin W-Y, Camp NJ, Cannon-Albright LA, Allen-Brady K, Balasubramanian S, Reed MW, Hopper JL, Apicella C, Giles GG, Southey MC, Milne RL, Perez JIA, Menéndez-Rodríguez P, Benítez J, Grundmann M, Dubrowinskaja N, Park-Simon T-W, Dörk T, Garcia-Closas M, Figueroa J, Sherman M, Lissowska J, Easton DF, Dunning AM, **Rajaraman P**, **Sigurdson AJ**, **Doody MM**, **Linet MS**, Pharoah PD, Schmidt MK, Cox A. A role for XRCC2 gene polymorphisms in breast cancer risk and survival. J Med Genet. 2011; 48(7):477-84.
85. **Linet MS**, Kim KP, Miller DL, **Kleinerman RA**, **Simon SL**, **de González AB**. Historical review of occupational exposures and cancer risks in medical radiation workers. Radiat Res. 2010; 174(6):793-808.
86. **Meinhold CL**, **Ron E**, **Schonfeld SJ**, Alexander BH, **Freedman DM**, **Linet M**, **Berrington de González A**. Non-radiation risk factors for thyroid cancer in the US Radiologic Technologists Study. Am J Epidemiol. 2010; 171(2):242-52.
87. Milne RL, Gaudet MM, Spurdle AB, Fasching PA, Couch FJ, Benitez J, Arias Perez JI, Zamora MP, Malats N, Dos Santos Silva I, Gibson LJ, Fletcher O, Johnson N, Anton-Culver H, Ziogas A, Figueroa J, Brinton L, Sherman ME, Lissowska J, Hopper JL, Dite GS, Apicella C, Southey MC, **Sigurdson AJ**, **Linet MS**, **Schonfeld SJ**, **Freedman DM**, Mannermaa A, Kosma VM, Kataja V, Auvinen P, Andrulis IL, Glendon G, Knight JA, Weerasooriya N, Cox A, Reed MW, Cross SS, Dunning AM, Ahmed S, Shah M, Brauch H, Ko YD, Bruning T, Genica Network T, Lambrechts D, Reumers J, Smeets A, Wang-Gohrke S, Hall P, Czene K, Liu J, Irwanto AK, Chenevix-Trench G, Holland H, Kconfab Investigators T, Aocs Investigators T, Giles GG, Severi G, Baglietto L, Bojesen SE, Nordestgaard BG, Flyger H, John EM, West DW, Whittemore AS, Vachon C, Olson JE, Fredericksen Z, Kosel M, Hein R, Vrieling A, Flesch-Janys D, Heinz J, Beckmann M, Heusinger K, Ekici AB, Haeberle L, Easton DF, Humphreys MK, Morrison J, Pharoah PD, Garcia-Closas M, Goode EL, Chang-Claude J. Assessing interactions between the associations of common genetic susceptibility variants, reproductive history and body mass index with breast cancer risk in the Breast Cancer Association Consortium: a combined case-control study. Breast Cancer Res. 2010; 12(6):R110. PMID:21194473
88. **Neta G**, **Brenner AV**, Sturgis EM, Pfeiffer RM, Hutchinson A, Aschebrook-Kilfoy B, Yeager M, Xu L, Wheeler W, Abend M, Ron E, Tucker MA, Chanock SJ, **Sigurdson AJ**. Common genetic variants related to genomic integrity and risk of papillary thyroid cancer. Carcinogenesis. 2011; 32(8):1231-7.
89. **Neta G**, **Rajaraman P**, **Berrington de Gonzalez A**, **Doody MM**, Alexander B, Preston D, **Simon S**, Melo D, Miller J, **Freedman DM**, **Linet MS**, **Sigurdson AJ**. A prospective study of medical diagnostic x-rays and risk of thyroid cancer. American Journal of Epidemiology 2012. [Epub ahead of print]; doi: 10.1093/aje/KWS315.
90. **Neta G**, Yu CL, **Brenner A**, Gu F, Hutchinson A, Pfeiffer R, Sturgis EM, Xu L, **Linet MS**, Alexander BH, Chanock S, **Sigurdson AJ**. Common genetic variants in the 8q24 region and risk of papillary thyroid cancer Laryngoscope. 2012; 122(5):1040-2.
91. Nickels S, Truong T, Hein R, Stevens K, Buck K, Behrens S, Eilber U, Schmidt M, Häberle L, Vrieling A, Gaudet M, Figueroa J, Schoof N, Spurdle A, Rudolph A,

Fasching P, Hopper J, Makalic E, Schmidt D, Southey M, Beckmann M, Ekici A, Fletcher O, Gibson L, dos Santos Silva I, Peto J, Humphreys M, Wang J, Cordina E, Menegaux F, Nordestgaard B, Bojesen S, Lanng C, Anton-Culver H, Ziogas A, Bernstein L, Clarke CA, Brenner H, Müller H, Arndt V, Stegmaier C, Brauch H, Brüning T, Harth V, The Genica Network; Mannermaa A, Kataja V, Kosma V-M, Hartikainen J, kConFab Investigators, AOCS Management Group, Lambrechts D, Smeets D, Neven P, Paridaens R, Flesch-Janys D, Obi N, Wang-Gohrke S, Couch F, Olson J, Vachon C, Giles G, Severi G, Baglietto L, Offit K, John E, Miron A, Andrulis I, Knight J, Glendon G, Mulligan AM, Chanock S, Lissowska J, Liu J, Cox A, Cramp H, Connley D, Balasubramanian S, Dunning A, Shah M, Trentham-Dietz A, Newcomb P, Titus L, Egan K, **Cahoon E**, **Rajaraman P**, **Sigurdson A**, **Doody M**, Guénel P, Pharoah P, Schmidt M, Hall P, Easton D, Garcia-Closas M, Milne R, Chang-Claude J. Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. PLOS Genetics. Submitted.

92. **Schonfeld SJ**, Bhatti P, Brown EE, **Linet MS**, **Simon SL**, Weinstock RM, Rosenstein M, Stovall M, Preston DL, Alexander BH, **Doody MM**, **Sigurdson AJ**. Polymorphisms in oxidative stress and inflammation pathway genes, low-dose ionizing radiation, and the risk of breast cancer among U.S. radiologic technologists. Cancer Causes Control. 2010; 21(11):1857-66.
93. **Schonfeld SJ**, Hartge PA, Pfeiffer RM, **Freedman DM**, Greenlee RT, **Linet MS**, Park Y, Schairer C, Schatzkin A, Visvanathan K, Lacey JV Jr. An aggregated analysis of hormonal factors and endometrial cancer risk. Cancer, in press.
94. **Schonfeld SJ**, **Neta G**, Sturgis EM, Pfeiffer RM, Hutchinson AA, Xu L, Wheeler W, Guénel P, **Rajaraman P**, de Vathaire F, **Ron E**, Tucker MA, Chanock SJ, **Sigurdson AJ**, **Brenner AV**. Common genetic variants in sex hormone pathway genes and papillary thyroid cancer risk Thyroid. 2012; 22(2):151-6.
95. **Schonfeld SJ**, Pfeiffer RM, Lacey JV Jr, **Berrington de González A**, **Doody MM**, Greenlee RT, Park Y, Schairer C, Schatzkin A, **Sigurdson AJ**, Hartge P, Visvanathan K. Hormone-related risk factors and postmenopausal breast cancer among nulliparous vs parous women: an aggregated study. Am J Epidemiol. 2011; 173(5):509-17.
96. **Sigurdson AJ**, Bhatti P, Chang SC, **Rajaraman P**, **Doody MM**, Bowen L, **Simon SL**, Weinstock RM, **Linet MS**, Rosenstein M, Stovall M, Alexander BH, Preston DL, Struewing JP. Polymorphisms in estrogen biosynthesis and metabolism-related genes, ionizing radiation exposure, and risk of breast cancer among US radiologic technologists. Breast Cancer Res Treat. 2009; 118(1):177-84.
97. **Simon SL**. Organ-specific external dose coefficients and protective apron transmission factors for historical dose reconstruction for medical personnel. Health Phys. 2011; 101(1):13-27.
98. Thierry-Chef I, **Simon SL**, Weinstock RM, Kwon D, **Linet MS**. Reconstruction of absorbed doses to fibroglandular tissue of the breast of women undergoing mammography (1960 to the present). Radiat Res. 2012; 177(1):92-108.
99. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, González -Bosquet J, Prokunina-Olsson L, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD,

Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver R, Prentice R, Jackson R, Kooperberg C, Chlebowski R, Lissowska J, Peplonska B, Brinton LA, **Sigurdson A**, **Doody M**, Bhatti P, Alexander BH, Buring J, Lee IM, Vatten LJ, Hveem K, Kumle M, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF, Jr., Hoover RN, Chanock SJ, Hunter DJ. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet. 2009; 41(5):579-84.

100. Yu CL, Falk RT, Kimlin MG, **Rajaraman P**, **Sigurdson AJ**, Horst RL, Cosentino LM, **Linet MS**, **Freedman DM**. The impact of delayed blood centrifuging, choice of collection tube, and type of assay on 25-hydroxyvitamin D concentrations. Cancer Causes Control. 2010; 21(4):643-8.
101. Yu CL, Li Y, **Freedman DM**, Fears TR, Kwok R, Chodick G, Alexander B, Kimlin MG, Kricker A, Armstrong BK, **Linet MS**. Assessment of lifetime cumulative sun exposure using a self-administered questionnaire: reliability of two approaches. Cancer Epidemiol Biomarkers Prev. 2009; 18(2):464-71.

3. Radiosensitive Organ Sites

BMI Pooling

102. **Berrington de González A**, Hartge P, Cerhan JR, Flint AJ, Hannan L, MacInnis RJ, Moore SC, Tobias GS, Anton-Culver H, Freeman LB, Beeson WL, Clipp SL, English DR, Folsom AR, **Freedman DM**, Giles G, Hakansson N, Henderson KD, Hoffman-Bolton J, Hoppin JA, Koenig KL, Lee IM, **Linet MS**, Park Y, Pocobelli G, Schatzkin A, Sesso HD, Weiderpass E, Willcox BJ, Wolk A, Zeleniuch-Jacquotte A, Willett WC, Thun MJ. Body-mass index and mortality among 1.46 million white adults. N Engl J Med. 2010 Dec 2; 363(23):2211-9.
103. **Kitahara CM**, **Neta G**, Pfeiffer RM, Kwon D, Xu L, Freedman ND, Hutchinson AA, Chanock SJ, Sturgis EM, **Sigurdson AJ**, **Brenner AV**. Common genetic variants in obesity and metabolism pathways and risk of papillary thyroid cancer. Cancer Epidemiol Biomarker Prev. In press..
104. **Kitahara CM**, Platz EA, Beane Freeman LE, Black A, Hsing AW, **Linet MS**, Park Y, Schairer C, **Berrington de González A**. Physical activity, diabetes, and thyroid cancer risk: a pooled analysis of five prospective studies Cancer Causes Control. 2012 Feb 1. [Epub ahead of print]
105. Moore S, Patel A, Matthews C, **Berrington de González A**, Park Y, Schatzkin A, **Linet M**, Weiderpass E, Visvanathan K, Helzlsouer KJ, Thun M, Hartge P, Lee I-M. Physical activity, body mass index and all cause mortality: a pooled analysis of 650,000 adults. PLOS Medicine. In press.

Glioma (Genetic) GWAS

106. Bhattacharjee S, **Rajaraman P**, Jacobs KB, and 40 others including **Inskip PD**. A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. Am J Hum Genet. 2012; 90:821-835.

107. Jacobs KB, Yeager M, Zhou W, and 186 others including **Rajaraman P, Inskip PD**. Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. 2012; 44:651-658.
108. **Kitahara CM**, Wang SS, Melin BS, Wang Z, Braganza M, **Inskip PD**, Albanes D, Andersson U, Beane Freeman LE, Buring JE, Carréon T, Feychting M, Gapstur SM, Gaziano JM, Giles GG, Hallmans G, Hankinson SE, Henriksson R, Hsing AW, Johansen C, **Linet MS**, McKean-Cowdin R, Michaud DS, Peters U, Purdue MP, Rothman N, Ruder AM, Sesso HD, Severi G, Shu XO, Stevens VL, Visvanathan K, Waters MA, White E, Wolk A, Zeleniuch-Jacquotte A, Zheng W, Hoover R, Fraumeni JF Jr, Chatterjee N, Yeager M, Chanock SJ, Hartge P, **Rajaraman P**. Association between adult height, genetic susceptibility, and risk of glioma. Int J Epidemiol 2012; 41:1075-1085.
109. Liu Y**,** Melin BS, **Rajaraman P,** Wang Z**, Linet M** et al. Insight into glioma susceptibility through an analysis of 6p22.3, 12p13.33-12.1, 17q22-23.2 and 18q23 SNP genotypes in familial and non-familial glioma. Hum Genet 2012; 131: 1507-17, epub.
110. Melin B, Dahlin AM, Andersson U, Wang Z, Henriksson R, Hallmans G, Bondy ML, Johansen C, Feychting M, Ahlbom A, **Kitahara CM**, Wang SS, Ruder AM, Carreón T, Butler MA, **Inskip PD**, Purdue M, Hsing AW, Mechanic L, Gillanders E, Yeager M, **Linet M**, Chanock SJ, Hartge P, **Rajaraman P**. Known glioma risk loci are associated with glioma with a family history of brain tumours – a case-control gene association study. Int J Cancer. In Press.
111. **Rajaraman P**, Melin BS, Wang Z, and 64 others including **Linet MS**, **Inskip PD**. Genome-wide association study of glioma and meta-analysis. Hum Genet 2012, Aug 11; Epub ahead of print.
112. **Rajaraman P**, Wang S, Melin B, Shete S, Kutsenko A, Carreon-Valencia T, Ruder AM, Feychting M, Johansen C, **Inskip PD**, **Linet M**, Beane-Freeman L, Hsing A, Safaeian M, Kraft P, Wolpin B, Bondy M, Chanock SJ, Hartge P, Yeager M. ABO blood group alleles and risk of glioma. Submitted.
113. Safaeian M, **Rajaraman P**, Hartge P, Yeager M, Butler MA, Ruder AM, Purdue MP, Hsing A, Beane-Freeman L, Hoppin JA, Albanes D, Weinstein SJ, **Inskip PD**, **Brenner A**, Rothman N, Chatterjee N, Gillanders EM, Chanock SJ, Wang SS. Joint effects between 5 identified risk variants, allergy and auto-immune conditions in glioma risk. Submitted.
114. Wang SS, Hartge P, Yeager M, Carréon-Valencia T, Ruder AV, **Linet M**, **Inskip PD**, Black A, Hsing A, Alavanja M, Beane-Freeman L, Safaiean M, Chanock SJ, **Rajaraman P**. Joint effects between 5 genetic variants and menstrual and reproductive factors in glioma risk among women. Am J Epidemiol. 2011; 174:901-908.

Second Breast Cancer

115. Nichols HB, **Berrington de González A**, Lacey JV Jr, Rosenberg PS, Anderson WF. Declining incidence of contralateral breast cancer in the United States from 1975 to 2006. J Clin Oncol. 2011; 29:1564-9.

4. Dosimetry

Dosimetry Methodologic Studies

116. DeWitt R, Klein DM, Eduardo G, Yukihara EG, **Simon SL**, McKeever SWS. Optically stimulated luminescence (OSL) of tooth enamel and its potential use in post-radiation exposure triage. Health Physics. 2010; 98(2): 432-439.
117. Sholom S, Desrosiers M, Chumak V, Luckyanov N, **Simon SL**, **Bouville A**. UV effects in tooth enamel and their possible application in EPR dosimetry with front teeth. Health Physics. 2010; 98(2): 360-368.
118. **Simon SL**, **Bouville A**, Kleinerman R. Current use and future needs of biodosimetry in studies of long-term health risk following radiation exposure. Health Physics. 2010; 98(2): 109-117.
119. **Simon SL**, Skinner AR, Swartz HL. Editor's remarks: biodose-2008 special issue. Health Physics. 2010; 98(2):93-94.

Kazakhstan Dosimetry

120. **Drozdovitch V**, **Schonfeld S**, Akimzhanov K, Aldyngurov D, **Bouville A**, **Land CE**, Luckyanov N, **Mabuchi K**, Potischman N, Schwerin MJ, Semenova Y, Tokaeva A, Zhumadilov Z, **Simon SL**. Behavior and food consumption pattern of the population exposed in 1949-1962 to fallout from Semipalatinsk nuclear test site in Kazakhstan. Radiation and Environmental Biophysics. 2011; 50(1):91-103.
121. Schwerin MJ, **Schonfeld S**, Potischman N, Zhumadilov Z, **Mabuchi K**, **Drozdovitch V**. Using focus group interviews to capture group-level estimates of dairy consumption in Kazakhstan villages in the 1950s. Journal of Developmental Origins of Health and Disease. 2010; (1)4:192-202. ERRATUM published 1(4):280, 2010.

Other Radiological Dosimetry Studies

122. Beck HL, **Bouville A**, Moroz BE and **Simon SL**. Fallout deposition in the Marshall Islands from Bikini and Enewetak nuclear weapons tests. Health Physics. 2010; 99(2): 124-142.
123. **Berrington de González A**, Iulian Apostoaei A, Veiga LH, **Rajaraman P**, Thomas BA, Owen Hoffman F, **Gilbert E**, **Land C**. RadRAT: a radiation risk assessment tool for lifetime cancer risk projection. J Radiol Prot. 2012 Sep;32(3):205-22.
124. **Bouville A**, Beck HL, **Simon SL**. Doses from external irradiation to Marshall Islanders from Bikini and Enewetak nuclear weapons tests. Health Physics. 2010; 99(2): 143-156.
125. Harris PS, **Simon SL**, Ibrahim, SA. Urinary Excretion of Radionuclides from Marshallese Exposed to Fallout from the 1954 Bravo Nuclear Test Health Physics. 2010; 99(2): 217-232;.

126. Ibrahim SA, **Simon SL**, **Bouville A**, Melo D, Beck HL. Alimentary tract absorption (f_1 Values) for radionuclides in local and regional fallout from nuclear tests. Health Physics. 2010; 99(2): 233-251.
127. **Land CE**, **Bouville A**, Apostoaei I, **Simon SL**. Projected lifetime cancer risks from exposure to regional radioactive fallout in the Marshall Islands. Health Physics. 2010; 99(2):201–215.
128. Moroz BE, Beck HL, **Bouville A**, **Simon SL**. Predictions of dispersion and deposition of fallout from nuclear testing using the NOAA-HYSPLIT meteorological model. Health Physics. 2010; 99(2): 252-269.
129. **Simon SL**, **Bouville A**, **Land CE**, Beck HL. Radiation doses and cancer Risks in the Marshall Islands associated with exposure to radioactive fallout from Bikini and Enewetak nuclear weapons tests: summary. Health Physics. 2010; 99(2): 105-124.
130. **Simon SL**, **Bouville A**, Melo D, Beck HL, Weinstock RM. Acute and chronic intakes of fallout radionuclides by Marshallese from nuclear weapons testing at Bikini and Enewetak and related internal radiation doses. Health Physics. 2010; 99(2): 157-200.

5. Other

Inter-lymph Consortium

131. Turner JJ, **Morton LM**, **Linet MS**, Clarke CA, Kadin ME, Vajdic CM, Monnereau A, Maynadié M, Chiu BC, Marcos-Gragera R, Costantini AS, Cerhan JR, Weisenburger DD. InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. Blood. 2010 Nov 18;116(20):e90-8. Epub 2010 Aug 10.

Retinoblastoma Follow- Up Study (Expansion Study)

Date Abstracted: |___|___|/|___|___|/|_2_|_0_|_0_|___|

Abstractor Initials: [ ]

Unique Patient Identifier: |__0_|__4_|___|___|___|___|___|

MSK Hospital Number: |___|___|___|___|___|___|___||___|

SOCIAL HISTORY

1. Date of Birth: |___||___|/|___||___|/|___||___||___||___|
   Month Day Year

2. Sex: |___|
   1 Male
   2 Female
   9 Not Mentioned

3. Race of Patient: |___|
   1 White
   2 Black or African American
   3 Asian or Pacific Islander
   4 American Indian or Alaska Native
   8 Other
   9 Unknown

4. Is patient Hispanic or Latino? |___|
   1 Hispanic or Latino
   2 Not Hispanic or Latino
   9 Unknown

5. Weight at time of first RB treatment: |___||___| lbs or |___||___||___| kg

6. Height at time of first RB treatment: |___||___| inches or |___||___||___| centimeters

7. Did patient have documented hearing loss prior to RB treatment? |___|
   0=No
   1=Yes
   7=Unclear/Discrepant Information
   9=Not Mentioned/Unknown

RETINOBLASTOMA (RB) DIAGNOSTIC HISTORY

8. RB Diagnostic Status:

   8a. Definitive RB Diagnosis made: |___|
   0=No-non-RB diagnosis (e.g., Coat's Disease) [STOP ABSTRACTING}
   1=Yes
   7=Unclear/Discrepant Information [STOP ABSTRACTING]
   9=Not Mentioned/Unknown [STOP ABSTRACTING]

8b. Study Center Involvement: |___|

0=Subject Never Seen – Consultation Only [STOP ABSTRACTING]
1=Subject Examined and Treated
2=Subject Examined Only – Referred Elsewhere for Treatment [CONSULT WITH SUPERVISOR}

9. Laterality of RB: |___|

1=Right Eye (OD)
2=Left Eye (OS)
3=Bilateral (OU)
4=Unilateral, Eye Unspecified
9=Unknown (STOP ABSTRACTING)

10. Was molecular testing for a RB germ line mutation performed?

10a. |___||___|

00=No
01=Yes
99=Unknown/Not Mentioned

10b. Photocopy enclosed: |___||___|

00=No
01=Yes
99=Unknown/Not Mentioned

Comments:

_____
_____
_____
_____
_____

11. **DESCRIPTION OF RB DIAGNOSIS(ES):**

Description	Right Eye (OD) or Eye Unspecified		Left Eye (OS)	
a. **Date of First Sign or Symptom** 99=Unknown/Not Mentioned	|___||___|/|___||___||___||___| Month Year		|___||___|/|___||___||___||___| Month Year	
b. **Date of First Diagnosis** 99=Unknown/Not Mentioned	|___||___|/|___||___||___||___| Month Year		|___||___|/|___||___||___||___| Month Year	
c. **Place First Diagnosed** Hospital/Physician Name: Address: Hospital Record #: 99999999=Unknown	_____ _____ _____ |___|___|___|___||___|___|___|___|		_____ _____ _____ |___|___|___|___||___|___|___|___|	
d. **Date First Seen at Study Center** 99=Unknown/Not Mentioned	|___||___|/|___||___||___||___| Month Year		|___||___|/|___||___||___||___| Month Year	
e. **Method of Diagnosis** Mark boxes for all methods of diagnosis that apply. *If method of diagnosis was by **CT scan**, please go to 10f. Otherwise, skip to 10g.*	☐ Enucleation ☐ Ocular biopsy/tapping Interior chamber ☐ Physical exam without obtaining tumor tissue ☐ Ultrasound ☐ MRI	☐ CT scan with contrast (go to f) ☐ CT scan without contrast (go to f) ☐ Unclear/discrepant ☐ Other(specify in comments) ☐ Unknown/Not mentioned	☐ Enucleation ☐ Ocular biopsy/tapping Interior chamber ☐ Physical exam without obtaining tumor tissue ☐ Ultrasound ☐ MRI	☐ CT scan with contrast (go to f) ☐ CT scan without contrast (go to f) ☐ Unclear/discrepant ☐ Other(specify in comments) ☐ Unknown/Not mentioned
f. **Number of CT scans performed:**	# of CT scans with contrast: |___| \# of CT scans without contrast: |___|		# of CT scans with contrast: |___| \# of CT scans without contrast: |___|	
g. **Surgical Pathology Report of Diagnosis, Photocopy Enclosed** 0=Not in Record 1=Yes	|___|		|___|	
h. **Number of Tumors** 01=One tumor 96=Multiple/Mass Tumors 97=Unclear/Discrepant 99=Unknown	|___||___|		|___||___|	
i. **Diagram of Eye, Photocopy Enclosed** |___| 0=Not in Record 1=Yes				
j. **Year Last Treated for RB (of any site)** 9999=Unknown/Not Mentioned	|___||___||___||___|		|___||___||___||___|	

Treatment for RB:

12. RADIATION FOR RB: |___|

Include **External Beam** and **Radiation Plaque** (Brachytherapy). Photocopy radiation records or summary of radiation treatment from medical progress notes in medical chart. **Please include any treatment for recurrent (relapsed) Retinoblastoma (RB)**.

0=No (If NO, skip to Question 13)
1=Yes
7=Unclear/Discrepant Information
9=Not Mentioned/Unknown

Radiation Therapy Table

Method 1=External 2=Brachy (Plaque) 9=Unknown	**Site Treated** 91=Right Eye (OD) 92=Left Eye (OS) 93=Both Eyes (OU) 94=Eye Unspecified 96=Non-ocular site (specify in comments)	**Date of First Treatment**	**Date of Last Treatment**	**Photocopy Enclosed** 0=Not in record 1=Yes
|___|	|___|___|	a. |___|___|/|___|___|___|___| Month Year	a. |___|___|/|___|___|___|___| Month Year	|___|
|___|	|___|___|	b. |___|___|/|___|___|___|___| Month Year	b. |___|___|/|___|___|___|___| Month Year	|___|
|___|	|___|___|	c. |___|___|/|___|___|___|___| Month Year	c. |___|___|/|___|___|___|___| Month Year	|___|
|___|	|___|___|	d. |___|___|/|___|___|___|___| Month Year	d. |___|___|/|___|___|___|___| Month Year	|___|
|___|	|___|___|	e. |___|___|/|___|___|___|___| Month Year	e. |___|___|/|___|___|___|___| Month Year	|___|

Comments:
_____
_____
_____
_____
_____

13. ENUCLEATION FOR RB:

Code status for each eye. If yes, record date of enucleation below.

0=No
1=Yes
7=Unclear/Discrepant Information
9=Not Mentioned/Unknown

a. Right Eye (OD) or Eye Unspecified		b. Left Eye (OS)	
Status	Month/Year	Status	Month/Year
|___|	|___|___|/|___|___|___|___| Month Year	|___|	|___|___|/|___|___|___|___| Month Year

14. CRYOTHERAPY, PHOTOCOAGULATION (LASER), AND THERMOTHERAPY FOR RB: |___|

If yes, record surgical treatment details below. Otherwise, skip to item 15. **Plaque placement is a surgical procedure but treatment data from plaque placement should be recorded in item 12, radiation treatment for RB.**

0=No (If NO, skip to Question 15)
1=Yes
7=Unclear/Discrepant Information
9=Not Mentioned/Unknown

Surgical Treatment Table

Surgery (Specify Type)	**Eye Treated** 1=Right Eye (OD) 2=Left Eye (OS) 3=Both Eyes (OU) 4=Eye Unspecified	**Date Treated** **Date Began** Month Year	**Date Ended** Month Year
a. _____	|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year
b. _____	|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year
c. _____	|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year
d. _____	|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year
e. _____	|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year
f. _____	|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year

Comments (reference surgical treatment line (a, b, c, etc) and record comments below):

_____
_____
_____
_____
_____

15. CHEMOTHERAPY FOR RB: |___|

If yes, record chemotherapy treatment details in table. Otherwise, skip to item 16. Photocopy chemotherapy treatment records.

0=No (If NO, skip to Question 16)
1=Yes
7=Unclear/Discrepant Information
9=Not Mentioned/Unknown

Chemotherapy Table

Chemotherapy Drugs	Dose	Number of Cycles	Date of First Treatment	Date of Last Treatment
a. _____ _____ _____	_____ _____ _____	|___|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year
b. _____ _____ _____	_____ _____ _____	|___|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year
c. _____ _____ _____	_____ _____ _____	|___|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year
d. _____ _____ _____	_____ _____ _____	|___|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year
e. _____ _____ _____	_____ _____ _____	|___|___|	|___|___|/|___|___|___|___| Month Year	|___|___|/|___|___|___|___| Month Year

Comments:
_____
_____
_____
_____
_____
_____

Neoplasms:

16. Malignant and Benign Neoplasms Other Than RB (for example, pineoblastoma, lipoma, meningioma):
If Yes, record details in table. Photocopy neoplasm report records (for example, pathology reports, diagnostic workups, surgery reports, clinical reports, etc..)

|___| 0=No (If NO, skip to Question 17)
1=Yes
7=Unclear/Discrepant
9=Unknown/Not Mentioned

Neoplasm Table

Diagnosis (Type/Site) Exclude RB metastasis or recurrences	**Date of Diagnosis** Month Year	**Place Diagnosed** (Hospital/Physician Address)	**Confirmation Source** (record source from table below)	**Photocopy of Confirmation Source(s)** 0=No 1=Yes
a. _____ _____ _____	|___||___|/|___||___||___||___| Month Year	_____ _____ _____ _____	|___| Specify:_____ _____ _____	|___|
b. _____ _____ _____	|___||___|/|___||___||___||___| Month Year	_____ _____ _____ _____	|___| Specify:_____ _____ _____	|___|

Confirmation Source:
1=Surgery Pathology Report
2=Autopsy Report
3=Radiology Report
4=Death Certificate
5=Clinical Report
6=Correspondence Only
7=Other (specify in table)
9=Unknown/Not Mentioned

Comments:
_____
_____
_____
_____
_____

Family History:

17. Family Pedigree, Photocopy Enclosed: |___|
    0=Not in Record
    1=Yes

18. Family History of RB: |___|
    0=No (If NO, skip to Question 19)
    1=Yes
    7=Unclear/Discrepant
    9=Unknown/Not Mentioned

    If Yes, complete table for each relative diagnosed with RB; otherwise skip to item 19.

Family History of RB Table

Relationship to Subject (Specify gender and maternal/paternal relationship)	**Laterality** 1=Unilateral 2=Bilateral 9=Unspecified	**Vital Status** 1=Alive 2=Deceased 9=Unknown
a. _____	|___|	|___|
b. _____	|___|	|___|
c. _____	|___|	|___|
d. _____	|___|	|___|
e. _____	|___|	|___|

Vital Status:

19. Date last seen at study center: |___||___|/|___||___|/|___||___||___||___|
    Month Day Year

20. Last known vital status and date: |___| 1 = alive (skip to Followup Information section)
    2 = deceased

    |___||___|/|___||___|/|___||___||___||___|
    Month Day Year

21. If Deceased, cause of death documentation, photocopy enclosed: |___|
    0=Not in record
    1=Yes
    7=Not applicable (subject alive)

22. If deceased, cause of death: _____

Comments:

_____

_____

_____

Followup Information:

Separate this page of the abstract form from the remainder of the form; it contains patient identifiers. Please store this page separately from the remainder of the abstract form and ship it separately from the other pages. Please record (print) parents' information below:

Father: Name: _____
Last First Middle Other

Social Security Number: **|___||___||___| - |___||___| - |___||___||___||___|**

Date of Birth: **|___||___|/|___||___|/|_1__||__9_||___||___|**
Month Day Year

Place of Birth:_____

Vital Status: |___| 1=Alive, 2=Deceased, 7=Unclear/Discrepant, 9=Unknown/Not Mentioned

Last Known Address:
_____
_____

Occupation:
_____
_____

Employer:
_____
_____

Mother: Name: _____
Last First Middle Other

Social Security Number: **|___||___||___| - |___||___| - |___||___||___||___|**

Date of Birth: **|___||___|/|___||___|/|_1__||__9_||___||___|**
Month Day Year

Place of Birth:_____

Vital Status: |___| 1=Alive, 2=Deceased, 7=Unclear/Discrepant, 9=Unknown/Not Mentioned

Last Known Address:
_____
_____

Occupation:
_____
_____

Employer:
_____
_____

Relatives/Friends/Neighbors	**Address Telephone Number**	**Year Noted**
Name:		

Comments:
_____
_____
_____

Field Administration

The medical record abstractor is the eyes and ears of the research team for this study. The function of the abstractor is to serve as a link between those who seek the data and the records that contain the medical information.

The abstractor's primary task is to collect specific data from medical records about selected patients. You, as an abstractor, are responsible for locating the correct records, then for verifying and extracting all pertinent information as accurately and completely as possible on to the abstract form.

You are valuable and necessary person on this research team. We view your role as one of a collaborator. This survey will only be as good as the abstractor(s) who work on it. Many private and public agencies have come to depend o accurate surveys to make decisions that affect people's health care; your job is especially important and meaningful.

Procedures for Data Collection

The abstraction effort for the Retinoblastoma (Rb) Follow-up Study (Expansion Study) involves abstracting information from the patient's full medical record onto a standard abstract form and obtaining follow-up information on the study subjects. These forms will be completed according to the item-by-item specifications outlined in this manual. It is crucial to follow the specifications as expressed in the manual so that the data will be collected in a standardized format and therefore, be comparable for analysis purposes.

The abstractor will receive a listing of the eligible study subjects (Rb patients) by patient name and by Memorial Sloan-Kettering Cancer Center (MSKCC) hospital number (if available) so that the medical records can be retrieved. In this study, the medical record to be abstracted should include the full patient chart, including all inpatient admissions and outpatient visits, the complete radiation treatment chart, and the hospital tumor registry file, starting from the date of first Rb diagnosis through the last date of available follow-up. The contractor will provide you with a set of labels pre-printed with unique identification numbers. You will assign an ID # to each eligible patient and affix a label with the appropriate ID number to each form, to each page of each form, and to all pages of photocopied records.

As the abstractor, you will be responsible for requesting and returning all records to the appropriate departments. It is best to wait until you have all the records available for a subject before you begin coding an abstract form. However, this may not be possible if only a small number of records are available. For example, records may be missing because the study subject is an active patient, either as an inpatient or an outpatient, or because the record may be signed out for review by someone else. Do not attempt to abstract incomplete information. Retain the abstract forms for these missing

records and periodically request the medical record again. If there is a long delay in accessing records, notify one of the study managers.

The hospital may have specific procedures for requesting, retaining, and returning medical records. These procedures will be discussed at the training session or will be communicated to you by staff in Dr. Abramson's office or by staff in collaborating departments. Take care when the records are in your possession that they are not lost, that they are kept in proper order, and that they are returned in the same condition as they were found.

Although you should have the patient's full medical record available to you, we ask you to abstract the record in a standard order. Begin with basic demographic information followed by the admission (in or outpatient) for the first diagnosis of Rb. Pay special attention to the medical record for all cancer related hospital admissions or visits. Other admissions or visits (non-cancer related) should be quickly reviewed to check for references to radiotherapy (radiation treatment), chemotherapy, or cancers. To accomplish this, the admission sheet and the discharge summary should be reviewed. End abstracting with the last admission or clinic visit so that you have collected data through the last available time. If the patient received radiotherapy as part of his/her treatment regimen, please visit the MSKCC radiation therapy department and try to obtain complete copies of the radiation therapy records including isotope, dose, specific anatomic location and schedule.

The abstracting is considered complete for each subject when the medical abstract and follow-up forms are completely coded and edited, when the photocopies are labeled and attached and when the completed forms are logged onto the appropriate reporting forms. Instructions for completing the reporting sheets and for sending data to the RTI office are presented in the next section. If you encounter any difficulties following the instructions given in this manual or at the training session, be sure to report them to the study manager. When questions arise about the procedures, notify the study manager, who will resolve problems and inform you of any procedural modifications.

Procedures for Conducting Quality Control

It is standard procedure that a specified percentage of all abstracting will be reviewed for quality control. For this study, quality control problems will involve the "blind" re-abstraction of the medical record by either the study manager or a peer abstractor. "Blind" reabstraction is done by a second abstractor who has not seen the first completed abstract. When the second abstraction is completed, the two forms can be compared for consistency in coding. It is preferable that the comparison of the editing is performed by the two abstractors at the hospital so that references to the medical record can be made. Alternatively, the comparison of the two completed forms will be performed by the study managers. If coding discrepancies are identified, problem sheets will be sent to the abstractor and the re-abstractor for discussion. Data relevant to resolving coding discrepancies will be retrieved by the abstractor(s) by referring to the medical records for the correct information and to the item-by-item specifications for the

correct coding conventions. This is an important step in the routine monitoring of the data collection and provides an opportunity for you to alert the home office about data collection problems so that procedures can be modified as necessary.

A 100% quality control check will be conducted at the initial stage of the study. When a satisfactory level of work is achieved, the number of abstracts to be redone for quality control will be reduced accordingly. The reabstraction for quality control is part of the routine schedule maintained throughout the data collection period as a continuing method to monitor completed work. In addition, all completed abstracts are reviewed at the home office by specially trained medical editors/coders and any coding problems are reported to the study manager.

Procedures Concerning Confidentiality

It is very important that, as researchers, we assure the participating hospitals that procedures will be followed to keep all data confidential.

Ethics of Survey Interviewing and Data Abstraction

Survey research is a profession where answers are sought to questions that may be considered as personal by the respondent. For the Rb Follow-up Study (Expansion Study), the subject's medical record from which you are abstracting data is not unlike a "respondent". Every hospital is concerned with maintaining privacy for each of its patients. Facilities are generally willing to allow abstraction of medical data if they feel that the information which is collected will be held in confidence. Thus, researchers must assure facilities that all data will be kept completely confidential. Only by making and keeping assurances of confidentiality can survey organizations succeed in collecting survey data.

To maintain confidentiality, each subject's data are combined with others, and the results are reported in such a way that individual data cannot be identified. It is your duty to keep the promise of confidentiality. Never divulge names or disclose facts about records you abstract. You should not discuss the contents of the medical record with co-workers in the presence of non-study staff. The information collected or observed in the course of the study can only be shared with the research team whose members are under the same ethical and moral obligation to preserve confidentiality as you are.

Confidentiality Legislation

The Public Health Service Act (42USC242m) Section 308(d) provides guidelines for safeguarding an individual's right to privacy. This Act requires that survey data collected for the federal government be kept absolutely confidential. To comply with these laws, RTI International ensures that information collected for this study will be gathered with a guarantee that it will be held in strict confidence, will be used only for the purposes stated in the study, and will not be disclosed or released to others without the consent of the individuals or facilities included in the study, except as required by law.

You should further understand that the collection of these data is authorized by the Public Health Service Act. Information collected under authority of the Public Health Survey Act may not be used for any purpose other than the purpose for which it is supplied. Also, such information may not be published or released in any form if the individual or the establishment is identifiable unless the individual or establishment has consented to such release.

The Privacy Act of 1974 (Public Law 93-579) is another piece of legislation concerned with the individual right to privacy. The Act ensures that all government studies provide safeguards against the invasion of personal privacy created by the collection, maintenance and use of personal information by Federal agencies. Since RTI International is under Federal contract to conduct the Rb Follow-up Study (Expansion Study), we are bound by the Privacy Act and the Public Health Service Act. Procedures have been developed for this study in accordance with these Acts. By complying with these procedures, you will ensure that data are collected in accordance with the law.

Confidentiality Procedures for This Study

RTI International is firmly committed to the principle that the privacy of individuals who provide data to us and the privacy of individuals and organizations about whom data are collected through survey methods must be protected. The contractor regards the confidentiality of records or other information regarding patients as being particularly important facets of the research process. Disclosure of any information is considered inconsistent with the rights of individual privacy.

Responsibility for confidentiality is assumed by the Radiation Epidemiology Branch of the National Cancer Institute, the contractor, by Memorial Sloan-Kettering Cancer Center, and by you, the abstractor.

RTI International will utilize a combination of established field and home office procedures to ensure that data collected in the course of the study that would reveal the identity of patients, the providers who care for them, or the facilities in which they are treated, will neither be disclosed nor made accessible to any individuals or organizations not connected with the study unless required by law.

The following confidentiality procedures have been developed specifically for this study:

- Medical records retained by the abstractor will be kept in locked cabinets or drawers, or locations specially designated by the hospital contact, and will be logged in/out of a field log as necessary to keep track of their use. Photocopies of medical record sheets will not leave the hospital until all personal identifiers are blacked out and the pages labeled by a unique study ID number.

- Study materials including patient lists, logs, and completed abstract forms at the site will be kept with the abstractor or in locked cabinets or drawers, or in a secure area as designated by the hospital contact, until shipped to RTI International.
- RTI International employees and consultants who are involved in collection or handling of the data will be thoroughly trained in matters pertaining to privacy and confidentiality.
- RTI International employees are required to sign a confidentiality statement affirming their responsibility to maintain confidentiality and their understanding of the sanctions to be imposed for violation (next page).
- Returned survey forms and other survey documents will be kept in locked files at RTI International with access limited to persons directly involved in the study.
- No data which can be linked with individual subjects will be transmitted to NCI or any other government or nongovernmental organization or person.
- No tabulations will be released in such detail that any subject can be identified by his or her personal characteristics or combination of characteristics.
- Strict control policies over tape and record duplication will be observed.

Source Documents for Retinoblastoma Expansion Study

The following documents should be obtained from medical records for each patient:

- Insurance frontsheet or other source document that lists demographic information for patient and his/her parents
- Medical progress notes detailing the Rb diagnosis and family history
- Surgery/exam/scan results used to diagnosis Rb
    - Enucleation surgery
    - Ocular biopsy
    - Physical exam
    - Ultrasound
    - MRI
    - CT scan
    - Pathology report
    - Placement of radiation plaque
- Radiation therapy records to document radiation type, method, dates of treatment, site treated
- Surgical records or notes documenting cryotherapy, photocoagulation (laser) and/or thermotherapy
- Chemotherapy treatment records documenting drug(s), dose, number of cycles, and dates of treatment
- Reports of neoplasms (malignant or benign) other than Rb

Copies should be made of the following documents:

- Eye diagram
- Pathology report from surgical (biopsy) procedure(s)
- Family pedigree
- Radiology report from scan used to diagnosis Rb
- Report detailing molecular testing for an Rb germline mutation (if applicable)
- Radiation therapy/treatment records
- Chemotherapy treatment records
- Reports of secondary neoplasms (benign or malignant)
- Death certificate/autopsy report

Specifications for completing each item from the Medical Record are given below:

Abstraction from Memorial Sloan Kettering Cancer Center Medical Records:

Administrative Information:

- **Date Abstracted:** Record the date the medical record was abstracted. This is the date the abstraction form is completed.
  Zero fill the month and day, and record the 4th digit for year (i.e., 07/01/2006).
- **Abstractor Initials:** Record the initials of the individual who is abstracting or reviewing the extracted cancer registry data.
- **Unique Patient Identifier:** Record the 7-digit patient identifier. This is a unique identification number assigned to each Rb patient:
  Site Number: digits 1-2 are pre-recorded as 04 (MSK location)
  Expansion Study: digit 3 is recorded as '3'
  Sequence Number: digits 4-7 are assigned to identify the individual patient
- **MSK Hospital Number:** Record the eight-digit MSK medical record number for the patient

Social History:

1. **Date of Birth:** Record the date of birth of the patient. Zero fill the month, day and record four digits for year (i.e., 07/07/2006). If after searching the patient's entire medical record and no date of birth is available, but the patient's age is recorded in the record AND this reference to age is dated, **it is acceptable to calculate the year of birth; do not calculate month or day of birth.** If birth year is calculated, record '99' for month and day, and record the calculated year of birth. Mark this item with an asterisk (*) and note your calculation in the COMMENTS field. For example, record the calculated date of birth as follows:

   |__9_||_9__|/|__9_||_9__|/|_1__||__9_||_8__||_8__|*
   Month Day Year

   Comments: *Patient was two years old at time of visit in 1990. 1990-2 = 1988 as the year of birth.

2. **Sex:** In the box, record the sex or gender of the patient. If the record does not state the patient's sex, you may attempt to determine if the patient is male or female by the use of pronouns in the record. However, be cautious. The medical record may have errors and could mislead you. Do not try to determine the patient's sex based on his/her name.
   1 = Male
   2 = Female

9 = Unknown (Record '9' if the patient's gender cannot be determined.)

3. **Race of Patient:** In the box, record the race of the patient. The record will usually include reference to race on the admission sheet, or in the admission history notes. You may ask the hospital contact person for assistance with race determination. Do not make a judgment regarding the patient's racial background based on the surname or place of birth.

   1 = White
   2 = Black or African American
   3 = Asian or Pacific Islander
   4 = American Indian or Alaska Native
   8 = Other
   9 = Unknown (Record '9' only if the patient's race cannot be determined.)

- An example of 'Asian or Pacific Islander' (4) would be Chinese, Filipino, Japanese, Korean, Vietnamese, Native Hawaiian, Guamanian or Samoan.
- Scandinavian, Jewish, Italian, etc. should not be coded as race; they indicate ethnicity.

Enter '9' if the patient's race is unknown or not mentioned, or if the information in the record is conflicting or unclear and race cannot be determined.

4. **Is Patient Hispanic or Latino?:** In the box, record if the patient is Hispanic or Latino.:

   1 = Hispanic or Latino
   2 = Not Hispanic or Latino
   9 = Unknown (Record '9' only if the patient's race cannot be determined.)

If the Hispanic or Latino ethnicity is unknown or not mentioned, or if the information in the record is conflicting or unclear, record '9'.

5. **Weight at time of first/initial/earliest Rb treatment:** Record the patient's weight (in pounds or kilograms) at the time of initial or first treatment for Rb. The record will usually include reference to weight prior to administering chemotherapy or radiation therapy – weight should be included on the administration form. If the patient received cryotherapy or photocoagulation (laser) treatment, review the medical progress notes just prior to treatment to try to determine weight. Please record the weight closest in time to the first treatment. If the date weight was determined is greater than two months prior to treatment, mark this item with an asterisk (*) and note the date that weight was obtained in the COMMENTS field. For Example:

Comments: *Date of weight is 05/07/04; treatment started on 7/8/04.

If the weight cannot be determined, enter "99" in the boxes for "lbs".

6. **Height at time of first Rb treatment:** Record the patient's height (in inches or centimeters) at the time of initial or first treatment for Rb. The record will usually include reference to height prior to administering chemotherapy or radiation therapy – height should be included on the administration form. If the patient received cryotherapy or photocoagulation (laser) treatment, review the medical progress notes just prior to treatment to try to determine height. Please record the height closest in time to the first treatment. If the date height was determined is greater than two months prior to treatment, mark this item with an asterisk (*) and note the date that height was obtained in the COMMENTS field. For Example:

Comments: *Date of height is 05/07/04; treatment started on 7/8/04.

If the height cannot be determined, enter "99" in the boxes for "inches".

7. **Did patient have documented hearing loss prior to Rb treatment?** If the patient received treatment for Rb, was there hearing loss documented prior to the first treatment? The record will usually include reference to sensory loss prior to treatment. Do not assume the patient had hearing loss if it is not documented. Only record '0' (No) for hearing loss if the record states that the patient had normal hearing or did not have hearing loss.

0 = No (patient did NOT have documented hearing loss)
1 = Yes (patient had documented hearing loss)
7 = Unclear/discrepant (hearing loss reported but timing is unclear)
9 = Not Mentioned/Unknown (not reported if the patient had documented hearing loss prior to treatment)

<u>Retinoblastoma (Rb) Diagnostic History:</u>

8. **Rb Diagnostic Status:**

8a. **Definitive Rb Diagnosis Made:** In the box, record if a definitive Rb diagnosis was made:

0 = No, non-Rb diagnosis (e.g., Coat's Disease) **[STOP ABSTRACTING]**
1 = Yes
7 = Unclear/Discrepant Information **[STOP ABSTRACTING]**
9 = Not Mentioned/Unknown **[STOP ABSTRACTING]**

8b. **Study Center Involvement:** In the box, record the medical status of the patient at Memorial Sloan Kettering Cancer Center (MSKCC):

0 = Subject Never Seen Consultation Only **[STOP ABSTRACTING]**
1 = Subject Examined and Treated
2 = Subject Examined Only – Referred Elsewhere for Treatment **[STOP ABSTRACTING]**

Record ‘0’ if the patient had a consultation only and was never actually seen at the study center. Sometimes, consultations of this nature are by correspondence. For example, the patient’s records, reports and or pathology slides may have been sent to the study center for review.

Enter code ‘1’ if the patient was examined and treated at the study center or an affiliated hospital, e.g., New York Hospital. Code “1” will be used for the Rb Expansion Study.

Record code ‘2’ if the patient was examined only at the study center and referred elsewhere for treatment. If treatment records are available, complete the abstract form. If treatment records are not included in the medical record, stop abstracting.

9. **Laterality of Rb:** In the box, record the laterality of the diagnosed Rb.

Note: This item is important in determining how to record data for Item 11. The type of laterality determined in item 9 should be consistent with the coding scheme for Item 11.

Right Eye = OD. OD = Oculus Dexter and is also abbreviated RE (right eye).
Left Eye = OS. OS = Oculus Sinister and is also abbreviated LE (left eye).
Both Eyes = OU. OU = Oculus Uterque.

Code ‘3’ applies whether the patient was diagnosed with bilateral Rb concurrently or at separate times.

1 = Right Eye (OD) (Rb in the right eye only)
2 = Left Eye (OS) (Rb in the left eye only)
3 = Bilateral (OU) (Rb in both eyes)
4 = Unilateral, Eye Unspecified (Rb in one eye but eye unspecified)
9 = Unknown (Record ‘9’ only if laterality cannot be determined.)

If an item in questions 1-10 commands further detail, asterisk the item and record your notes in the COMMENTS section after question 10.

10. **Was molecular testing for a Rb germline mutation performed?**

10a. In the box , record if molecular (genetic or gene) testing for an Rb germline mutation was performed:

Record code '00' (No) if the record specifies that molecular testing was not performed.

Record code '01' (Yes) if molecular testing was performed.

Record code '99' (Unknown/Not Mentioned) if the record does not indicate if molecular testing was performed.

10b. In the box, record if a photocopy of the molecular testing result(s) is available:
Record code '00' (No) if you are unable to photocopy the molecular testing result. If you are unable to photocopy the molecular testing results, specify in the COMMENTS.
Record '01' (Yes) if a photocopy of the molecular testing result is obtained and included in the submitted records. Please label the copy(ies) with the subject identification number, and attach it to the abstract form.

11. **Description of Rb Diagnosis(es):**
There are ten sub-categories (11a-j) of Item 11 displayed in table format that describe the patient's Rb diagnosis(es). Please note that the table layout is based on laterality (right or left) of the Rb diagnosis. Each sub-category must be coded under the appropriate heading (left or right eye) as applicable. The first heading (second column) is for "right eye or eye unspecified"). The second heading (third column) is for "left eye". ** The only exception to coding under the appropriate category/column is item 11i (diagram of eye, photocopy enclosed). In item 11i, record whether or not the photocopy is enclosed in column one under the description of Rb diagnosis(es).

If the laterality of Rb is right eye (coded as '1' in item 5) or it is unilateral but the laterality is unknown (coded '4' in item 5); the column "**RIGHT EYE (OD) OR EYE UNSPECIFIED**" must be completed.

If the laterality of Rb is left eye (coded as '2' in item 5), the column "**LEFT EYE (OS)"** must be completed.

If the laterality of Rb is bilateral (coded a '3' in item 5), both columns should be completed.

a. **Date of First Sign or Symptom:**

It is important to record the date of the first sign or symptom of retinoblastoma for each affected eye. Most often this is reported by the parents, but could include anyone who first noticed a dysfunction. For example, a common first sign is a report of "cat's eye" or "glassy eye".

Enter the date as month and year. If any portion of the date is not mentioned or is unknown, record as '99' for month and '9999' for year. Zero fill the month and record the four digit year (i.e., 07/2006).

b. **Date of First Diagnosis:**

For each affected eye, enter the date of the **first** Rb diagnosis made by a physician. Keep in mind that the first diagnosis may not have been at Memorial Sloan Kettering Cancer Center. Review the records carefully and record the date when the first Rb diagnosis was made. Record the date as month and year. If any portion of the date is not mentioned or is unknown, record as '99' for month and '9999' for year. Zero fill the month and record the four digit year (i.e., 07/2006).

c. **Place First Diagnosed:**

For each affected eye, enter the name and address of the hospital and/or physician where the patient was diagnosed with Rb, if available. Enter the hospital record number into the eight boxes provided. Left justify the numbers and leave remaining boxes blank.

Record '9's' in the boxes when the hospital record number is unknown. For convenience, you may enter a '9' in the first an last coding box with a line through the boxes between the two, e.g., "9—9".

When the hospital or physician at first diagnosis is the study center, MSK, or one of the study center physicians, you need only record the name and abbreviate the address, e.g., Dr. Abramson at MSK. Otherwise, complete both name and address for place first diagnosed.

d. **Date First Seen at Study Center:**

For each affected eye, record the date the patient was **first** seen at the study center, MSK or an affiliate hospital. Record the date as month and year. If any portion of the date is not mentioned or is unknown, record as '99' for month and '9999' for year. Zero fill the month and record the four digit year (i.e., 07/2006).

e. **Most Definitive Method of Diagnosis:**

This item refers to the method of diagnosis (or methods of diagnoses) used to confirm the Rb diagnosis. Careful review of the patient record is necessary. Mark all of the boxes that pertain to methods used to diagnosis Rb.

Mark the box for “Eunucleation” if the Rb diagnosis was made by enucleation. Enucleation is removal of the eye, whole and clean.

Mark the box for “Ocular biopsy/Tapping interior chamber” if the Rb diagnosis was made by an ocular biopsy, including tapping the anterior chamber. If any mention is made in the record of a **specimen being taken** and sent to pathology, mark this box.

Mark the box for “Physical exam without obtaining tumor tissue” if the Rb diagnosis was made by any exam where a tumor specimen was not taken. For example if the record indicates “exam under anesthesia” but there is no mention of a specimen being taken, mark this box.

Mark the box for “Ultrasound (sonogram)” if the Rb diagnosis was made following ultrasound or sonogram imaging. An ultrasound is a procedure in which high-energy sound waves (ultrasound) are bounced off internal tissues or organs and make echoes. The echoes form a picture of body tissues called a sonogram.

Mark the box for “MRI” is the Rb diagnosis was made following magnetic resonance imaging (MRI). This is a procedure that uses a magnet, radio waves, and a computer to make a series of detailed pictures of areas inside the body, such as the eye. This procedure is also called nuclear magnetic resonance imaging (NMRI).

Mark the box for “CT scan with contrast” if the Rb diagnosis was made following a CT scan (CAT scan) with contrast (or dye). This is a procedure that makes a series of detailed pictures of areas inside the body, such as the eye, taken from different angles. The pictures are made by a computer linked to an x-ray machine. This option includes using a dye that is injected into a vein or swallowed to help the organs or tissues show up more clearly. This procedure is also called computed tomography, computerized tomography, or computerized axial tomography.

Mark the box for “CT scan without contrast” if the Rb diagnosis was made following a CT scan (CAT scan) without contrast. This is a procedure that makes a series of detailed pictures of areas inside the body, such as the eye, taken from different angles. The pictures are made by a computer linked to an x-ray machine. This procedure is also called computed tomography, computerized tomography, or computerized axial tomography.

Mark the box for “unclear/discrepant” if there is discrepant or conflicting information in the record regarding the method(s) of diagnosis. Asterisk the item and record your notes in the COMMENTS.

Mark the box for “other (specify in comments)” if the Rb diagnosis was made by another method and specify the method of diagnosis in the COMMENTS.

Mark the box for "unknown/not mentioned" if the record makes no mention of the method if diagnosis or when the method of diagnosis is unknown.

f. **Number of CT scans performed:**

If the method of diagnosis in 11e is either "**CT scan with contrast**" or "**CT scan without contrast**" please record the number of CT scans performed **with** and **without** contrast during the diagnostic period. Careful review of the patient record is necessary to obtain the numbers of scans performed.

g. **Surgical Pathology Report of Diagnosis, Photocopy Enclosed**.

The record may contain a surgical pathology report of the patient's Rb diagnosis(es). We are interested in the summary report only of the surgical pathology report.

Record code '0' (not in record) only when a surgical pathology report is not included in the record and thus you are unable to make a photocopy.

Record code '1' (Yes) if a surgical pathology report is available in the patient record. Photocopy the summary page of this report and label the copy appropriately with the subject's study identification number.

If, for some reason, the surgical pathology report is in the medical record but you are unable to make a photocopy, record this information in COMMENTS at the end of Question 11.

h. **Number of Tumors:**

For each affected eye, enter the appropriate code as listed on the abstract form. The purpose of this item is to determine whether only one tumor was identified or more than one (multiple) tumors were found at diagnosis.

Record code '01' if the record indicates only one tumor.
Record code '96' if the record indicates multiple tumors or uses terms such as "masses" or "cottage cheese". Any reference to number of tumors as "multiple" or as "mass", e.g., "a mass of tumor" or "solid mass" should be coded as '96'.

Record code '97' if the record is unclear or discrepant about the number of tumors identified. Specify in COMMENTS.

Record code '99' when there is no mention in the record of the number of tumors diagnosed or it is unknown.

i. **Diagram of Eye, Photocopy Enclosed:**

The record may contain a diagram of the patient's eye at Rb diagnosis(es). We are interested in obtaining a photocopy of this eye diagram.

Record code '0' (not in record) only when a diagram of the eye is not included in the record and thus you are unable to make a photocopy.

Record code '1' (Yes) if a diagram of the eye is available in the patient record. Photocopy this diagram of the eye and label the copy appropriately with the subject's study identification number.

If, for some reason, the diagram of the eye is in the medical record but you are unable to make a photocopy, record this information in COMMENTS at the end of Question 10 and leave the box blank.

j. **Year Last Treated for Rb (of any site):**

This item refers to the last year the patient received treatment for Rb of any site. This includes treatment for recurrence and/or metastases of Rb. Do not include date of last exam if there was no treatment.

For each affected eye, record the four digit year the patient was last known to have received treatment for Rb (e.g., 2006).

If the date is not mentioned or is unknown, record as '9999'. This includes reference in the record of further treatment, but there is no confirming documentation that treatment was given.

Treatment for Rb:

Item 12 is concerned with radiation therapy (XRT) for Rb only, **which could include treatment to non-ocular sites of metastases**. *If the patient's weight and height at the time of radiation therapy is available, record this measurement in items 5 and 6 of this form. If the weight and height are different than their weight and height at time of first Rb treatment, choose the weight and height that are closest to the weight and height at the time of diagnosis. Note whether weight is given in pounds or kilograms. Do not convert from pounds to kilograms or vice versa. Note whether height is given in inches or centimeters. Do not convert from inches to centimeters or vice versa.

Include all treatment for Rb including treatment for recurrent (Rb that has come back after it has been treated) or relapsed Rb.

The patient's record may contain summary information regarding radiation treatment, but the summary information does not provide all the details that are requested to accurately determine the total dose absorbed by various organs of interest for this study. In addition, the information in the medical record may not include all the treatments actually received because some treatments reports never get filed in the

medical record. **Therefore, at the end of the abstraction, check all of the patient IDs against the files in the radiation treatment department to ensure that you have provided all of the pertinent radiation treatment records.**

Radiation therapy consists of **either** Brachytherapy (e.g., radium or radiation plaque) or external beam therapy, or both. Each of these types is described below:

Brachytherapy is internal radiation therapy and uses a radioactive substance sealed in needles, seeds wires, plaques, or catheters that are placed directly into or near the cancer. It may also be called plaque radiotherapy. Radioactive seeds are attached to one side of a disk, called a plaque, and placed directly on the outside wall of the eye near the tumor. The side of the plaque with the seeds on it faces the eyeball, aiming radiation at the tumor. The plaque helps prevent other nearby tissue from being exposed to the radiation.

Brachytherapy procedures usually are performed in an operating room since they require anesthesia. Therefore, in addition to radiotherapy records, some information about treatment may be found in the surgical procedure(s) notes in the record. **Although the Brachytherapy treatment may be performed in a surgical suite and recorded in surgical notes, it is radiotherapy and should not be considered surgery for purposes of describing the patient's treatment.**

External beam therapy is delivered by a machine which generates x-rays or contains a large amount of a radioactive isotope outside the body to send radiation toward the cancer. One or more beams of radiation are directed at specific anatomical sites so that the site receives the prescribed dose of radiation. External beam treatments are given in one or more "series" or "courses". Each course of radiation is administered over a period of days or weeks in small daily doses. The basic external beam record usually consists of a diagram showing the locations of the beams on the patient, a daily treatment log, and a table or notes which summarize the total dose delivered and other treatment parameters. This record may or may not be present in the patient's clinical file record.

Note: This item is concerned with radiation therapy NOT diagnostic x-rays (e.g., skull x-rays taken to help diagnose a condition).

Record code '0' (No) if the record clearly states that the patient did **not** receive radiation therapy (Brachytherapy or external beam therapy) for Rb. Do not complete the table and skip to item13.

Record code '1' (Yes) if the patient received radiation for Rb. Again, treatment site could be the eye or any non-ocular site of metastases, e.g., the orbital bone. Complete the Radiation Therapy Table.

Record code '7' (Unclear/Discrepant Information) if the record contains confusing or inconsistent information regarding radiation therapy. For example, the record may

state that radiation therapy was advised but no documentation is in the record. Do not complete the table, skip to item 13. Asterisk the item and record your notes in the COMMENTS. Photocopy clinic notes or references regarding plan for radiation treatment if possible.

Record code '9' (Not Mentioned/Unknown) if radiation therapy is not mentioned in the patient record. Do not complete the table; skip to item 13.

Radiation Therapy Table

If item 12 is coded "1" (Yes) radiation treatment was given, you should complete the Radiation Therapy Table. Instructions for completing the table are as follows:

- Make a separate entry for each treatment course. By course, we mean a series of doses or radiation over a prescribed period. For example, a patient was given five series of radiation doses over a six-month period. This would constitute **one** course of treatment.

- If the radiation record is available from the Radiation Treatment Department, photocopy all pages of the radiation record. Include radiation treatment for all Rb diagnoses including recurrent or relapsed Rb.

Note: Remember to review the patient's clinic record and to record and photocopy any additional information regarding radiation treatment that may not be included in the radiotherapy record (e.g. subject received additional radiation treatment at another institution).

- **Method** – Code the method of radiation therapy for each treatment course for Rb.

Record code '1'when the method was external beam (external).

Record code '2' when the method was Brachytherapy (brachy). Brachytherapy for Rb will usually have been administered as a plaque. **Do not attempt to specify which isotope (e.g., Ruthenium) was used** – the photocopied document/note should provide this information or a plaque identification number. However, it is still necessary to record the information on the abstract form for "site treated". **If the radiation treatment was not performed at MSK, record the hospital/physician name and address in COMMENTS.**

Record code '9' when the method of radiation therapy is not mentioned in the medical record yet it's clear that the patient received radiation treatment.

- **Site Treated** – Use the following codes as specified:

  Code '91' when the right eye (OD) only was treated;

Code ‘92’ when the left eye (OS) only was treated

Code ‘93’ when both eyes (OU) were treated;

Code ‘94’ when it is not specified as to which eye received radiation treatment;

Code ‘96’ when treatment was given for Rb to a non-ocular site. Specify the non-ocular site in COMMENTS (e.g., a non-ocular site would be the brain if Rb had metastasized to the brain).

Note: If the radiation treatment was not given at MSK, record the name and address of the hospital, the physician, or radiologist where treatment was given in the COMMENTS after Question 11.

● **Date of First Treatment** – Record the month and year the radiation treatment (implantation or courses) began for a particular course. Record ‘99’ for any portion of the date that is unknown.

● **Date of Last Treatment** – Record the month the year the radiation treatment was completed for a particular course. Record ‘99’ for any portion of the date that is unknown.

● **Photocopy Enclosed** – For each entry, code whether or not a photocopy of the radiation records is enclosed. **A photocopy should be made of the complete radiotherapy record whenever available**. **Remember to photocopy any document/notes pertaining to radiation treatment when the radiotherapy record is not available**. If possible, slightly reduce the copy so that all margins are included; this assures the dosimetrist that nothing is lost at the top or bottom of the page or the left or right margins of the record. If dates and/or doses are difficult to read on the photocopies, write the information on the photocopy either above or below the illegible information and trace over lines that diagram the body or treatment fields. **These records are absolutely essential for study success.**

Record ‘0’ (Not in record) if the radiotherapy record or other documentation regarding the treatment course is not available in the record. Specify in COMMENTS.

Record ‘1’ (Yes) if a photocopy of the radiotherapy record or other documentation of the treatment course is enclosed. Label the copy(ies) appropriately with the subject identification number and attach it to the abstract form.

If, for some reason, the diagram of the eye is in the medical record but you are unable to make a photocopy, record this information in COMMENTS at the end of Question 11 and leave the box blank.

13. **Enucleation for Rb:**

Enucleation, removal of the eye whole and clear, is the most common form of surgery for retinoblastoma. The patient record will often refer to enucleation in it's abbreviated form, "enuc".

Code the appropriate status for each eye. If the patient has unilateral Rb, do not code the date for the unaffected eye.

Record code '0' (No) when the patient record clearly states that the patient had **no** enucleation for Rb or when it is **evident** from the patient record that the eye(s) in question was examined **at the last visit**.

Record code '1' (Yes) when the patient had enucleation for Rb. Any mention of the term "socket" indicates that an enucleation was performed. For each affected eye, record the date as month and year when enucleation was performed. For any portion of the date that is unknown, code '99' or '9999' as appropriate.

Record code '7' (Unclear/Discrepant Information) when the record contains confusing or inconsistent information regarding enucleation. Record the date as indicated. If the date of enucleation is discrepant, try to make a determination as to which date may be more accurate and enter that date in item 13. Place an asterisk (*) at item 13 and record your notes in COMMENTS.

Record code '9' (Not Mentioned/Unknown) when enucleation for Rb is not mentioned in the record or when it is unknown if the patient had enucleation performed. Code the date as '99/9999'.

14. **Cryotherapy, Photocoagulation (laser) and Thermotherapy for Rb:**

Item 14 is concerned with types of surgical treatment for Rb **other than enucleation**. It is not uncommon to have more than one type of surgery or numerous courses of the same type of surgery performed on a particular eye. Cryotherapy, photocoagulation and thermotherapy are three common types of surgery for Rb.

Cryotherapy is a treatment that uses an instrument to freeze and destroy abnormal tissue such as a tumor. This type of treatment is also called cryosurgery.

Photocoagulation is a procedure that uses laser light to destroy blood vessels to the tumor causing tumor cells to die. Photocoagulation may be used to treat small tumors. This method is also called light coagulation or laser surgery.

Thermotherapy is the use of heat to destroy cancer cells. Thermotherapy may be given using a laser beam aimed through the dilated pupil or onto the outside of the eyeball, or using ultrasound, microwaves, or infrared radiation (light that cannot be seen but can be felt as heat).

Record code ‘0’ (No) when the record clearly indicates the patient had **no** cryotherapy, photocoagulation, or other types of surgical treatment for Rb. Do not complete the table; go to item 15.

Record code ‘1’ (Yes) when the patient record indicates that cryotherapy, photocoagulation, or other types of surgical treatment was performed. Complete the Surgical Treatment Table.

Record code ‘7’ (Unclear/Discrepant Information) when the patient record contains confusing or inconsistent information regarding surgical treatment for Rb. Place an asterisk (*) at item 14 and record your notes in COMMENTS.

Record code ‘9’ (Not Mentioned/Unknown) when the record does not mention or when it is unknown if the patient had any cryotherapy, photocoagulation or other surgical treatment for Rb. Skip the treatment table; go to item 15.

Surgical Treatment Table

If item 14 is coded as ‘1’ (Yes), the patient had surgical treatment, you must code the Surgical Treatment Table. Instructions for completing the table are as follows:

- **Surgery** – Make an entry for each type of surgery performed. Record the type of surgery performed; cryotherapy, photocoagulation, or other. **More than one course of treatment for a given type of surgery should be entered as one type**.

- **Eye Treated** – Code for the eye treated with respect to the type of surgery performed:
  - Code ‘1’ when the right eye (OD) only was treated;
  - Code ‘2’ when the left eye (OS) only was treated;
  - Code ‘3’ when both eyes (OU) were treated;
  - Code ‘4’ when eye is unspecified as to type of surgical treatment.

- **Date Treated** – For each type of surgery, record the date treatment first began and the date treatment ended. Numerous courses of a specific type of treatment may cover a span of years. **The “date began” would be the month and year the first course of treatment started and the “date ended” would be the month and year the last course of treatment ended.** Record the date as month and four-digit year. For any portion of the date that is unknown, code ‘99’.

For example, in various sections of a patient record, cryotherapy treatment was reported as given on five separate occasions to the right eye. The patient received cryotherapy treatment on April 6, 1991; September 12, 1991; January 3, 1992; June

20, 1992, and again on March 7, 1993. Item 14 would be coded ‘1’ (Yes); you would write “cryotherapy” on Line a of the Surgical Treatment Table. You would record ‘1’ for Right Eye for Eye Treated, record Date Began as 04/1991; and record date ended as 03/1993.

If you need to record comments, such as discrepant treatment information, please record this text in the COMMENTS section following Question 14.

15. **Chemotherapy for Rb**:

Chemotherapy is a cancer treatment that uses drugs to stop the growth of cancer cells, either by killing the cells or by stopping the cells from dividing. A type of chemotherapy called **chemoreduction** is used to treat retinoblastoma. **Chemoreduction reduces the size of the tumor so it may be treated with local treatment such as radiation therapy, cryotherapy, photocoagulation, or thermotherapy**. Chemotherapy may be the primary treatment, prescribed in cases of advanced cancer, in the hope or prolonging survival or relieving pain. Chemotherapy may be an adjuvant treatment, given in addition to surgery and/or radiation therapy, in the hope of curing patients at risk of recurrent cancer.

The patient record may or may not contain chemotherapy treatment data. Unlike surgery and radiation treatments, which must be performed at a hospital or specially equipped clinic, chemotherapy may be administered at a physician’s office. In addition, it may be self-administered under the guidance and program supervised by a physician. Therefore, it is essential that you search the patient record for any reference to chemotherapy. **A protocol may be comprised of a single chemotherapy agent/drug or a combination of agents administered over a designated period**. Each course of chemotherapy, whether administered as a single agent or a combination of agents, should be recorded only once.

For example, one treatment for Rb is Carboplatin (CDDP) via peri-ocular administration. Some other systemic chemotherapeutic agents include cytoxan, vincristine, cyclophosphamide and adriamycin. These agents have been administered singularly or in combination. Refer to the Appendix A for a listing of chemotherapy agents.

Record code ‘0’ (No) if the patient record clearly states that the patient did not receive chemotherapy for Rb. Do not complete the Chemotherapy Table; skip to item 16.

Record code ‘1’ (Yes) if the record states that the patient received chemotherapy treatment for Rb. Remember that a chemotherapy treatment protocol can consist of one agent or a combination of agents given over a designated period. Complete the Chemotherapy Table.

Record code ‘7’ (Unclear/Discrepant Information) if the record contains confusing or inconsistent information regarding chemotherapy for Rb. Place an asterisk (*) and record your note in the COMMENTS. For example, if the record indicates that chemotherapy was recommended, considered, or given, but no other documentation exists to support administration of the chemotherapy, code as ‘7’ and record your notes in the COMMENTS. Do not complete the Chemotherapy Table; skip to item 16.

Record code ‘9’ (Not Mentioned/Unknown) if the record does not mention chemotherapy for Rb or it is unknown from the patient record that the subject received chemotherapy. Do not complete the Chemotherapy Table; skip to item 16.

Chemotherapy Table

If item 15 is coded as ‘1’ (Yes), patient received chemotherapy, you must complete the Chemotherapy Table. Instructions for completing the table are as follows:

- **Chemotherapy Drugs** – Make a separate entry for each treatment protocol. In this table, you will document the history of each chemotherapy protocol agent or combination of agents given. **A specific protocol may have been given more than once; however, it is still considered the same protocol and only one entry should be recorded for that protocol.** Please record the name of each chemotherapy drug/agent given to the patient. If a protocol reference is available and you can identify the drugs, please list the drugs rather than the protocol number. Record the name(s) of the chemotherapy drugs/agents as recorded in the medical record. If the method of administration is not intravenous (IV), please document the route of administration in the COMMENTS section; other routes include periocular (literally within the eye; i.e., CDDP/Cisplatin) and intra-arterial (a method to administer chemotherapy directly to an artery or an organ. The medication travels through a catheter that is attached to a pump. The pump may be portable or implanted; i.e., Melphalan).

- **Dose** – Record the dose of each chemotherapy administered if this information is available in the medical record.

- **Number of Cycles** – Record the number of cycles (courses) for each specific chemotherapy administered.

- **Date of First Treatment** - Record the month and year treatment was first given for a specific protocol. Record ‘99’ for any portion of the date that is unknown.

- **Date of Last Treatment** – Record the month and year treatment was last given for a specific protocol. Remember a specific protocol may include numerous courses or cycles over a span of years. Be sure to record the date of the last course of treatment for that protocol (or chemotherapy drug(s)) as “date of last treatment”. Record ‘99’ for any portion of the date that is unknown.

- Use the COMMENTS field at the bottom of the table if there are additional entries for chemotherapy treatment.

Neoplasms:

16. **Malignant and Benign Neoplasms Other Than Rb**

The term "neoplasm" means a new and abnormal formation of tissue or a tumor or growth. It serves no useful function but grows at the expense of the healthy organ.

Item 16 is concerned with both **malignant and benign neoplasms**. A malignant neoplasm or cancer is a growth that infiltrates healthy tissue, continues to grow at a rapid rate, and is capable of metastasizing (spreading to other sites in the body). A benign neoplasm or benign tumor is a growth that does not spread by metastases or infiltration of tissue; it is considered non-cancerous. An example is a **lipoma**.

Review the patient record very carefully for any mention of a cancer (other than the Retinoblastoma diagnosis). **Include references to lipoma, preleukemia and to precancerous changes such as cysts**. When referring to breast neoplasms, always note whether the right, left, or both breasts are affected.

Do not list metastases or recurrent cancer; only list incident or new reports of cancer. Do not list Rb recurrences (Rb that was treated but came back) or relapsed Rb. However, whenever you are uncertain whether or not a neoplasm is a metastases or a recurrence, record the information from the patient record. Record as code '1'.

Record code '0' (No) if the patient record clearly states that the patient did not develop any malignant or benign neoplasms other than Rb. Skip to item 17.

Record code '1' (Yes) to indicate that the patient has been diagnosed with a neoplasm. Do not record metastases or recurrences. Complete the Neoplasm Table.

Record code '7' (Unclear/Discrepant Information) when the patient record contains confusing or conflicting information regarding neoplasms. Place as asterisk (*) next to the item and record your notes in the COMMENTS. For example, if the patient record indicates the patient had an examination for a brain tumor; however, no documentation is available in the record, record '7'. Skip to item 17.

Record code '9' (Unknown/Not Mentioned) when the patient record does not mention neoplasm development, or if the record states that development of a neoplasm is unknown. Do not complete the Neoplasm Table; skip to item 17.

Neoplasm Table

If item 16 is coded as '1' (Yes), you must complete the Neoplasm Table. Instructions for completing the table are as follows:

- **Diagnosis (Type/Site)** – Record the neoplasm diagnosed including the type and site, if available. **Record verbatim from the patient record for malignant and benign neoplasms**. If additional space is required to record the diagnosis, please asterisk (*) the item and record your notes in COMMENTS.

Note: It is extremely important that diagnostic information be as specific as possible. Examples of specific pediatric diagnoses are:

Type	Site
osteosarcoma	left leg
rhabdomyosarcoma	head
meningioma (benign)	brain
lipoma (benign)	upper right quadrant on back

For **leukemias** the **type** of leukemia is important, e.g., acute, chronic, myeloid, lymphatic, etc. There will probably NOT be a site listed for leukemia. Record only what is recorded in the record.

Type	Site
acute myeloid leukemia	none
chronic lymphocytic leukemia	none

- **Date of Diagnosis** – Record the month and year of the **earliest** confirmed diagnosis for each neoplasm. For any portion of the date that is unknown, code '99'.

- **Place Diagnosed** – Record the name and address of the hospital or physician where the neoplasm diagnosis was made. When the hospital that diagnosed the malignant or benign neoplasm is the study center, MSK, or one of the study center physicians, you need only record the name and abbreviate the address, e.g., Dr. Abramson at MSK. Otherwise, complete both name and address for place that the neoplasm was diagnosed.

- **Confirmation Source** – Use codes 1-9 (as recorded in the box below) to indicate the most reliable confirmation source of the neoplasm. Use code '9' only when the confirmation source is unknown.

Codes:

1 = Surgical Pathology Report
2 = Autopsy Summary Report
3 = Radiology Report
4 = Death Certificate
5 = Clinical Report only

6 = Correspondence only
7 = Other (specify in table)
9 = Unknown/Not Mentioned

Note: If you code '7', for "Other" be sure to specify the confirmation source in the table or record your note in COMMENTS.

Some confirmation types are preferable to others. **The confirmation sources are listed in order of priority, i.e., the lower the code the higher the priority**. For example, if a neoplasm was confirmed by an autopsy summary report and a radiology report, the correct code to record would be '2' (autopsy summary report). Although the coding of "confirmation source" is based on the best documentation available, you are expected to review all sources documenting the neoplasm.

- **Photocopy of Confirmation Source**

Complete documentation of the neoplasm is essential to this study. You are expected to provide photocopies of **every** source which documents the neoplasm. This includes pathology reports, diagnostic work-ups, clinical reports, etc. Specify sources copied in COMMENTS.

Record '0' (No) if you are unable to make a photocopy documenting diagnosis of the neoplasm. Asterisk (*) the item and specify why you were unable to obtain a photocopy in COMMENTS.

Record '1' (Yes) if a photocopy is enclosed documenting diagnosis of the neoplasm. Indicate sources photocopied in COMMENTS.

Note: If the patient was diagnosed with more than two neoplasms, mark an asterisk (*) beside item 15 and record details in COMMENTS.

Family History

17. **Family Pedigree, Photocopy Enclosed**

Some patient records may contain a family pedigree or a family tree, which traces the lineal pattern of hereditary retinoblastoma disease within a family. A photocopy of the family pedigree is required whenever it is available in the patient's record.

Record '0' (Not in Record) when a family pedigree is not available in the patient record so you are unable to make a photocopy. Specify your notes in COMMENTS.

Record '1' (Yes) when a family pedigree is available in the patient record and a photocopy is enclosed.

If, for some reason, the family pedigree is in the medical record but you are unable to make a photocopy, record this information in COMMENTS at the end of Question 22.

18. **Family History of Rb:**

When the patient record indicates a family history of Rb, record this information in item 18. **Rb is a hereditary disease and information on family history is vital to the study.**

Record '0' (No) when the record clearly states that there is no family history of Rb. Do not complete the Family History of Rb Table; skip to item 19.

Record '1' (Yes) if the record indicates that there is a family history of Rb. Complete the Family History of Rb Table. Instructions for completing the table are below.

Record '7' (Unclear/Discrepant) when the record contains confusing or inconsistent information regarding the family history of Rb. Place an asterisk (*) by item 18 and record your notes in COMMENTS.

Record '9' (Unknown/Not Mentioned) when the record does not mention a family history of Rb or states that it is unknown. Do not complete the Family History of Rb table; skip to item 19.

Family History of Rb Table

- Record an entry for each relative diagnosed with Rb. Include all biological relatives; distant or not, e.g., mother, father, grandfather, cousin, maternal aunt.

- **Relationship to Subject** – Specify the relationship of the relative to the subject. Where applicable, specify all relatives as **M=maternal or P=paternal** and their gender. Refer to the listing of relationships below:

Mother
Father
Brother
Sister
Aunt (Maternal/Paternal)
Uncle (Maternal/Paternal)
Cousin (Maternal/Paternal, gender)
Grandmother (Maternal/Paternal)
Grandfather (Maternal/Paternal)

- **Laterality** – Record the laterality of the Rb diagnosis for each relative:
  Record '1'= Unilateral (only one eye affected)
  Record '2'= Bilateral (both eyes affected)
  Record '9' = Unspecified (Unknown/Not Mentioned)

- **Vital Status** – Record vital status for each relative listed in the table. Record '1' when the most recent information indicates the relative is alive. Record '2' when the record indicates the relative is deceased. Code '9' when no mention of vital status is made in the record.

Vital Status

19. **Date last seen at study center** – Record the month, day, and year the patient was last seen by a physician at the study center. For this Rb Expansion study, the study center may be Memorial Sloan Kettering or NY Hospital. Record '99' for any portion of the date that is unknown.

20. **Last Known Vital Status and Date** – Record the last known vital status of the patient and the date of the last known vital status as noted in the record. Code the vital status as follows:

- Code '1' = Alive (skip to Follow-Up Information)
- Code '2' = Deceased

Record the date of the last known vital status as month, day, and year. For deceased patients, this will be the date of death. For living patients, this will be the last date they were known to be alive either through a visit or other contact.

Please note that we are interested in the actual date of death or date last known living and **not** the date such information was reported to the study center.

Record '99' for any portion of the date that is unknown.

21. **Cause of Death Documentation, Photocopy Enclosed**

Item 20 requests documentation of cause of death. For deceased patients, we are interested in obtaining all known causes of death. The best sources for obtaining this information are the death certificate and the autopsy summary report. If neither are available, any notes in the patient record documenting cause of death should be photocopied.

Record '0' (Not in Record) when no documentation is available in the record.

Record '1' (Yes) when either the death certificate, the autopsy summary report, clinical notes or physician letter(s) are available in the patient record. When any of all of the following are available, make a photocopy and attach to the abstract form. Both the death certificate and autopsy summary report are preferable when available. If neither are available, photocopy the clinical notes.

Record ‘2’ (No) when death documentation is available in the patient record but you are unable to make a photocopy. Asterisk (*) and record your notes in COMMENTS.

Record ‘7’ (Not applicable – subject alive) when the patient’s vital status is alive.

22. **If deceased, cause of death**

If the patient is deceased and you are able to locate a cause(s) of death, copy the text verbatim from the record. If the cause of death is not available from the record, death certificate or autopsy summary report, record “Unknown”.

Follow-Up Information Sheet Specifications

A Follow-up Information Sheet must be completed for each patient in the study. The information collected on this sheet will be used to find a current address for a living patient or to find a death certificate for a deceased patient. Please note that the Follow-up Information Sheet will eventually be separated from the rest of the abstract form. Be sure that you have recorded the Subject Identification Number at the bottom of each page, and any overlapping material (e.g., study identification number, vital status, date of birth, etc.) on both forms. This sheet gives the **ONLY** information the follow-up team will have for locating the patient. In some cases, many years will have elapsed between the last contact and the present time. **The more information you can locate in the medical record the easier their job will be**.

Be sure that each item is completed. Use the convention "NA" to indicate "Not applicable", and "NM" to indicate "Not mentioned" or "Unknown". For example, employer is not applicable (NA) for a stay-at-home mother. **Always record the most recent information available and note the latest year for which it is confirmed**.

The amount of information that must be collected for each patient depends on the last known vital status of the patient and other factors. Please carefully follow the instructions in the remained of this section.

For patients who are confirmed or assumed to be alive, and for patients who are confirmed to be deceased, complete all items. Collect as much information as possible. Be sure to include known relatives/friends/neighbors including previous husbands. Include all telephone numbers whenever available. **A good source for parent names and social security number is insurance forms**. Often, the most current contacts, e.g., family physician, insurance company, institution, can provide the best tracing information. Carefully complete this information as requested at the end of the Follow-up Information Sheet. A patient is assumed to be alive if s/he cannot be confirmed as deceased (e.g., death certificate, autopsy report, etc.).

APPENDIX A

Chemotherapy Drugs:

Carboplatin –Paraplatin, CBDCA

Cisplatin - CDDP

Cyclophosphamide – CTX, CPM, Cytoxan, Endoxan

Cyclosporine A – CSA, Sandimmune

Dactinomycin – ACT, Actinomycin-D, Cosmegen

Doxorubicin – ADR, Advarmycin

Etoposide – VePesid, VP-16

Fluorouracil – FU, Adrucil, 5-FU, 5-Fluorouracil, Fluoroplex

Melphalan – Alkeran, L-Pam, L-Sarcolysin, Phenylalanine Mustard

Methotrexate – MTX, A-methopterin, Sodium Methotrexate

Mitomycin – Mitomycin-C, Mutamycin

Vincristine – Vincristine Sulfate, VCR, Oncovin

APPENDIX B

Medical Terms

Aqueous humor
Watery fluid which bathes and nourishes the front of the eye.

Bilateral retinoblastoma
Cancerous tumor(s) in the retina of both eyes.

CAT Scan
Test which uses x-rays to view the eyes and brain.

Choroid
A layer of blood vessels that feeds the retina.

Chromosome 13
The chromosome which has a missing piece that is responsible for the development of retinoblastoma.

Conjunctiva
Thin membrane which lines the outside of the eye.

Cornea
The clear outer part of the eye's focusing system light entering the eye.

Cryotherapy
Freezing treatment for small retinoblastoma tumor(s).

Disc Diameter
Horizontal size of the optic nerve head approximately 1-1.5 mm used as a reference point to measure tumors.

Enucleation
Surgical removal of the eye.

Equator
A circular reference zone approximately half way between the back of the eye and the front of the eye.

External Beam Radiation
Treatment which uses machines to give radiation to treat the tumors.

Eyelid
The skin-covered structure that protects the front of the eye.

Fovea
The center of the macula that gives the sharpest vision.

Fundus Drawing
Map of the eye with tumor sketches drawn by the ophthalmologist.

Indirect Ophthalmoscope
Instrument used to view the retina.

Iris
The colored part of the eye.

Lamina Cribrosa
The zone in the optic nerve which represents the anatomical end of the eye.

Laser
Light therapy used to treat small retinoblastoma tumors(s).

Lens
The clear part of the eye behind the iris that helps to focus light on the retina.

Leukocoria
Which pupillary reflex; cats' eye reflex; the most common presenting sign of retinoblastoma.

Macula
The small sensitive area of retina that gives central vision.

Metastasis
Spread of a cancer to other parts of the body.

Optic Nerve
The bundle of over one million nerve fibers that carries visual messages from the retina to the brain.

Ora Serrata
Portion of the retina near the front part of the eye.

Pediatric Oncologist
Physician who specializes in cancer of children.

Prognosis
The overall outlook of treatment.

Pupil
The opening at the center of the iris that controls the amount of light.

Radioactive I-125 Plaque
A disc of radiation that is sewn to the eye to treat the retinoblastoma tumor(s).

Retina
The light-sensitive tissue lining the back of the eyeball.

Retinoblastoma
A cancerous tumor of the retina of the eye.

Sclera
The tough, white outer coat of the eye.

Strabismus
Crossed eyes; second most common presenting sign of retinoblastoma.

Ultrasound
Test which uses sound waves to view the eye and tumor(s).

Unilateral retinoblastoma
Cancerous tumor(s) in the retina of one eye.

Vitreous Humor
The clear gel filling the inside of the eye.

Vitreous Seeding
Small balls of retinoblastoma that break off and extend into the vitreous.

APPENDIX C

Pediatric Oncology Abbreviations

This section aims to provide a reference list of abbreviations and acronyms that are frequently used within oncology. As well as many widely used medical abbreviations this text includes abbreviations that have meaning only within oncology, sub-specialities or even particular organisations. No definitive list of standard abbreviations exists therefore it is important to understand the context in which they have been used in. For example "PNET" usually refers to a type of brain tumour, however, within orthopaedic oncology the same abbreviation has been applied to a form of sarcoma. You may also find it useful to have access to a medical dictionary or text-book that contains a list of common abbreviations used within the wider field of medicine.

This list of has been compiled over a number of years and abbreviations have been collected from a wide range of sources including journal articles and conference presentations.

Acknowledgements Thanks to the following who have contributed additional abbreviations to this list; Jane Evans, Kevin Fishwick, Anna Mitchell, Moira Stewart and the UKCCSG Data Centre.

Abbreviations

5-FU 5-Fluorouracil (anti cancer drug)
6-MP 6-Mercaptopurine (anti cancer drug)
6-TG 6-Thioguanine (anti cancer drug)
AA Anaplastic Anaemia
ABMT Autologous bone marrow transplant
ADR Adverse Drug Reaction
AE Adverse event
AFP Alphafetoprotein - eg. expressed by germ cell tumours and other cancers
AIDS Acquired immune deficiency syndrome
ALAT Alanine aminotransferase / alinine transaminase
ALCL Anaplastic Large-cell Lymphoma
ALL Acute lymphoblastic leukaemia
ALT Alanine Aminotransferase
AMKL acute megakaryocytic leukemia
AML Acute Myeloid leukaemia
ANC Absolute neutrophil count
ANED Alive no evidence of disease
ANLL Acute non-lymphatic leukaemia

ARMS	Alveolar rhabdomyosarcoma
ASR	Age Standardised Rate (Incidence)
AUC	Area under the curve
B-ALL	B-cell Acute Lymphoblastic Leukaemia
BAER	Brainstem Auditory Evoked Responce
BCC	Basal Cell Carcinoma
BID / BD	Twice a day (bis in die)
BM	Bone Marrow
BM	Blood Monitoring (eg for glucose)
BMR	Basal Metabolic Rate
BMT	Bone Marrow Transplant
BP	Blood pressure
BRM	Biological Response Modifier
BSA	Body Surface Area
Bx	Biopsy
C/O	Complaining of
C/W	Continue With
C1 - C7	Cervical vertebrae (spine eg. C7 = seventh cervical vertebra)
Ca	Cancer; carcinoma
Ca	Calcium
cALL	Common ALL
CAT	Computerised axial tomography (scan)
cc	Cubic centimeter
CCR	Continuous complete remission
CGH	Comparative Genomic Hybridisation - cytogenetics method
CGL	Chronic Granulocytic Leukaemia
cGy	Centi Gray (unit of radiation)
CLL	Chronic lymphocytic Leukaemia
cm	centimeter - 0.01 meters
CML	Chronic myeloid leukaemia
CMML	chronic myelomonocytic leukemia
CMV	Cytomegalo virus
CNS	Central nervous system - the brain and spine
CPM	Cyclophosphamide (anti cancer drug)
CPR	Cardio pulmonary resuscitation
CR	Complete remission / complete response
CRF	Case Report Forms

CSF	Cerebro spinal fluid
CSF	Colony-stimulating Factor
CT	Computerised axial tumography (scan)
CT	Chemotherapy
CTC	Common Toxicity Criteria
CTCL	Cutaneous T-Cell Lymphoma
CVC	Central venous catheters
CVP	Central Venous Pressure
CXR	Chest X-Ray
D/C	Discharge
D/H	Drug History
D/W	Discussed With
DDx	Differential diagnosis
DFI	Disease Free Interval
DFS	Disease Free Survival - time without disease prior to relapse or last follow-up
DI	Diabetes Incipidus
dl	deciletre - 0.01 litres
DLBCL	Diffuse Large B-cell Lymphoma
DLCL	Diffuse large-cell lymphoma
DLS	Date last seen
DLT	Dose limiting toxicity - determined by phase 1 studies
DMC	Data Monitoring Commitee
DNA	Deoxyribonucleic acid
DNA	Did Not Attend (clinic)
DNR	Do Not Resusitate
Dx	Diagnosis
EBV	Epstein-Barr Virus
ECG	Electrocardiogram - heart scan
EDTA	ethylendiaminetetraacetic acid - used in measuring kidney function
EEG	Electroencephalogram - brain scan
EFS	Event Free Survival - time from diagnosis to defined events (eg relapse or deat
ENT	Ear nose throat
ESR	Erythrocyte Sedimentation rate
ETS	Environmental Tobacco Smoke
F/H	Family history
FAB	French American and British classification scheme for leukaemia
FBC	Full Blood Count
FEV	Forced expectorant volume (a lung test)

FFA	For Further Appointment
FISH	Flourescence in situ Hybridisation
FNA	Fine Needle Aspiration - a type of biposy using a thin needle (or FNAB)
FU	Follow up
FVC	Forced Vital Capacity
g	gram - unit of weight
G-CSF	Granulocyte colony stimulating factor promotes production of white blood cells
GA	General Anaesthetic
GCP	Good Clinical Practice (guidelines)
GCT	Germ Cell Tumour
GFR	Gromerular filtration rate
GI	Gastrointestinal
GM-CSF	Granulocyte and macrophage colony stimulating factor
GPR	Good Partial Remission
GU	Genito-urinary
GvHD	Graft versus Host Disease
Gy	Grays (units of radiation)
H&E	Hematoxylin and Eosin (stain)
H/O	History of
Hb	Haemoglobin
HCL	Hairy Cell Leukaemia
HCO3	Bicarbonate
HD	Hodgkin's Disease (lymphoma)
HD	High dose
HDC	High Dose Chemotherapy
HIV	Human Immunodeficiency Virus
HL-A	Human Leukocyte Associated antigens (HL-A matching for BMT)
HR	High risk
I-131	Radioactive Iodine
ICD	International Classification of Diseases (coding system)
ICDO	International Classification of Diseases for Oncology (coding system)
ICF	Intercellular fluid
ICU	Intensive Care Unit
IM	Intramuscular - into a muscle
IMRT	Intensity-Modulated Radiotherapy
IU	International units

IV	Intravenous - into a vein
K+	Potassium
kg	Kilogram - a thousand grams
l	liter - unit of volume
L1 - L5	Lumbar vertebrae 1 - 5 (spine eg. L1 = 1st lumbar vertebra)
LN	Lymph Node
LP	Lumbar puncture
m	meter (unit of length)
M/H	Medical history
MDR	Multi drug resistant
mEq/l	milliequivalent per liter
mets	Metastases (where the tumour has spread to secondary sites)
Mg	Magnesium
mg	milligram - 0.001 gram
ml	millilitre 0.001 liter
mM	millimole
mm	millimeter - 0.001 meters
mOsm	milliosmole
MRI	Magnetic resonance imaging (scan)
MTX	Methotrexate (anti cancer drug)
MUD	Matched Urelated Donor - for bone marrow transplant
N/V	Nausea and vomiting
Na+	Sodium
NAD	No Abnormality Detected
NBM	Nil by mouth
NED	No evidence of disease
ng	nanogram - 0.000000001 gram
NHL	Non Hodgkin's Lymphoma
NK	Natural Killer cells (large lymphocytes, part of the immune system)
NK	Not known
NMR	Nuclear magnetic resonance (scan)
NOS	Not otherwise specified (see ICDO)
NSR	Non significant result
O/E	On Examination
OS	Osteogenic sarcoma (context bone tumours)
PBSC	Peripheral Blood Stem Cell (see PBSCT)
PBSCH	Peripheral Blood Stem Cell Harvest
PBSCR	Peripheral Blood Stem Cell Rescue (transplant)

PBSCT	Peripheral Blood Stem Cell Transplant
PD	Progressive disease
PET	Positron Emmission Tomography- a scan after a small radioactive injection.
pg	picogram - 0.000000000001 gram
pH	hydrogen-ion concentration - acid / alkaline
PH	Past History
PR	Partial Responce / Partial Remission
prn	as required
prn	whenever necessary (pro re nata)
qid	Four times a day (quater in die)
QoL	Quality of Life
RBC	Red blood cell / red blood count
RCT	Randomised Controlled Trial
RFS	Relapse free survival - Time from diagnosis to relapse or death.
RIGS	Radioimmunoguided surgery
RMS	Rhabdomyosarcoma
RT	Radiotherapy
Rx	Treatment
SA	Surface area (see BSA)
SAE	Serious Adverse Event
SC	Subcutaneous
SD	Stable Disease
SH	Social history
SOB	Short of breath
TBI	Total body irradiation
TCP	Thrombocytopenia
tds / tid	Three times a day (ter in die)
TNM	Staging system - primary tumour
UA	Urine analysis
ug	microgram - 0.000001 gram
ULN	Upper Limits of Normal
URTI	Upper respiratory tract infection
US	Ultasound (scan)
UVR	Ultra Violet Radiation
VM-26	Teniposide (anti cancer drug)
VMA	Vanillylmandelic Acid
VP-16	Etoposide (anti cancer drug)
WBC	White blood cell count

WCC White cell count
XRT Radiotherapy (external)
YST Yolk sac tumour - (aka. germ cell tumour)

APPENDIX D

Medical Record Abstraction

Procedures for Chart Abstraction:

MSK employee will abstract medical records for approximately 350 Rb patients using abstraction procedures for each data item documented in a Manual of Operations.

Medical Record Abstraction:

Training:

- Project Manager will train abstractor (MRA) at MSK:
    - ❖ Project Manager will train MRA using 3 historical Rb patient charts with de-identified data as source material. Using the Manual of Operations, the Project Manager will train the MRA to abstract each variable on the abstraction form.
    - ❖ Following the initial abstraction, the MRA will complete 2 test abstractions using the remaining 2 historical charts; the Project Manager and the abstractor will discuss any issues that arise during the abstraction.
    - ❖ Subsequently, the MRA will abstract 3 medical records for newly identified Rb patients using hard copy medical charts and the MSK computer system. The Project Manager will review these abstracted forms with the MRA.
    - ❖ PM will revise the Manual of Operations as needed during the abstraction phase of the study.

Quality Assurance:

- Project Manager will re-abstract the first 25 records completed by the MRA.

- Project Manager will re-abstract 10% of the remaining records (assume 325 records remain; 10% = 33).

Subject ID Assignment and Selecting Records for Re-abstraction:

- ❖ RTI will generate and provide MRA with a spreadsheet containing Subject ID numbers and a place for recording Subject Name.
- ❖ MRA will assign each subject an ID number. When the MRA pulls each medical chart for abstraction, she will assign each patient name to a subject ID number on the master log (*refer to Master ID Log spreadsheet).* She will record the subject name assigned to each subject ID number.
- ❖ MRA will abstract 25 records put them aside for review. PM will travel to MSK, re-abstract the records, and retrain the MRA as needed.
- ❖ RTI will re-abstract these records under a HIPAA Waiver for this activity.
- ❖ MRA will continue to abstract the remaining 325 medical records.
- ❖ PM will re-abstract 10% of the remaining records (assume 325 records remain; 10% = 33).
- ❖ Subsequent to abstracting the first 25 records, the MRA will send RTI an electronic list of the "unassigned" subject ID numbers.
- ❖ A programmer/analyst will use a SAS function to assign random numbers to the remaining 325 subject ID numbers. She will "sort" the random numbers and the PM will select records 1-33 for re-abstraction. . If additional records require QA, the PM will re-abstract additional medical records from the sorted list.
- ❖ PM will travel to MSK to monitor abstraction during two site visits; one site visit will occur about a month after the start of abstraction; the second will occur about one month after the first trip.
- ❖

Storage at MSK:

- MRA will store the first 25 completed MRA forms and source documents, by ascending subject ID number, for QA by the PM.
- MRA will store the first 8 pages of each MRA Form separately from the last page (page 9 contains identifying data).
- Following re-abstraction of the first 25 records, for each subject with a signed HIPAA Authorization Form (*refer to Tracking spreadsheet)*, the MSK Research Assistant (RA) will copy each completed MRA Form and source documents to retain for their records and send the original MRA Form and source document to RTI-Rockville for review and editing. Pages 1-8 will be shipped separately from page 9.

Transmission to RTI:

- Following QA of the first 25 records, the RA will ascertain from the tracking spreadsheet (maintained by RTI) whether or not HIPAA Authorization Forms have been received for the completed MRA Forms.
- For selected MRA Forms (with a signed HIPAA Authorization), MSK will ship pages 1-8 of each abstraction form separately from page 9, the last page, to RTI-Rockville. PM will review and edit the forms as necessary. PM will re-abstract selected records (n=~33) as described above.
- PM will FedEx the edited MRA forms to the Fulfillment Department in RTP for keying (pages 1-8 will be shipped and keyed separately).
- At the end of the abstraction phase, the MRA forms without corresponding, signed HIPAA Authorization Forms will be retained at MSK and not released to RTI.

PARENT QUESTIONNAIRE

Subject Identifier: |__0_|__4_|___|___|___|___|___|

Date: |___|___|/|___|___|/|_2_|_0_|_0_|___|

Relationship to patient of person completing form: _____

Demographic Information and Health History

1. When was your child born? |___|___|/|___|___|/|___|___||___|___|

2. In what city and state was your child born?
_____

3a. Our records indicate that your child was diagnosed with the eye tumor called retinoblastoma (Rb) between 1996 and 2006. Not including examinations, has your child been treated for Rb **since** then?

|___| 0=No (go to question 4a)
1=Yes (go to question 3b)
9 = Don't Know (go to question 4a)

3b. If your child has been treated since then, (yes to item 3a), please complete the treatment information below:

Treatment for Retinoblastoma

Type of Treatment (Surgery, Chemotherapy, Radiation)	**Date Treatment Began (Month/Year)**	**Place of Treatment (Hospital/Physician Name and Address)**
	|___||___|/|___||___||___||___| Month Year	
	|___||___|/|___||___||___||___| Month Year	
	|___||___|/|___||___||___||___| Month Year	

4a. Has your child been diagnosed with hearing loss since Rb treatment?

|___| 0=No (go to question 5a)
1=Yes (go to question 4b)
9 = Don't Know (go to question 5a)

4b. If yes, where did hearing loss occur?

|___| 1 = Right Ear
2 = Left Ear
3 = Both Right and Left Ears
9 = Don't Know

4c. If yes, what was your child's age when hearing loss was diagnosed?

_____ months or _____ years

5a. The following questions refer to any tumors (cancers) that your child may have had – **other than** retinoblastoma. Has your child ever been examined or treated for other tumors or cancers; this may have included a biopsy, surgery, radiation treatment or chemotherapy?

|___| 0=No (go to question 6)
1=Yes (go to question 5b)
9 = Don't Know (go to question 6)

5b. If your child has been examined for or treated for other tumors or cancers (yes to item 5a), please complete the information below for each tumor/cancer:

Treatment for Other Tumors/Cancers

Cancer or Tumor Diagnosed (Specify type and site)	Type of Treatment (chemotherapy, radiation, biopsy, other)	Date of Diagnosis (Month/Year)	Place of Diagnosis (Hospital/Physician Name and Address)
		|___||___|/|___||___||___||___| Month Year	
		|___||___|/|___||___||___||___| Month Year	
		|___||___|/|___||___||___||___| Month Year	

6. I am required by law to give you the following information:

The National Institutes of Health is requesting your child's social security number under Section 301 of the Public Health Service Act. His/her social security number may be used to link your child with vital statistics records. Disclosure of your child's social security number is voluntary, and your child will not be denied any Federal right, benefit, or privilege by your refusal to disclose it.

What is your child's social security number? **|___||___||___|-|___||___|-|___||___||___||___|**

7. In case we wish to contact you again in the future, it would be a great help to us if you could provide us with the names and addresses of one or two people who could give us your new address should you move. We would only contact these persons if we were unable to reach you at your home address, and we would not disclose any personal health information that you have disclosed to us.

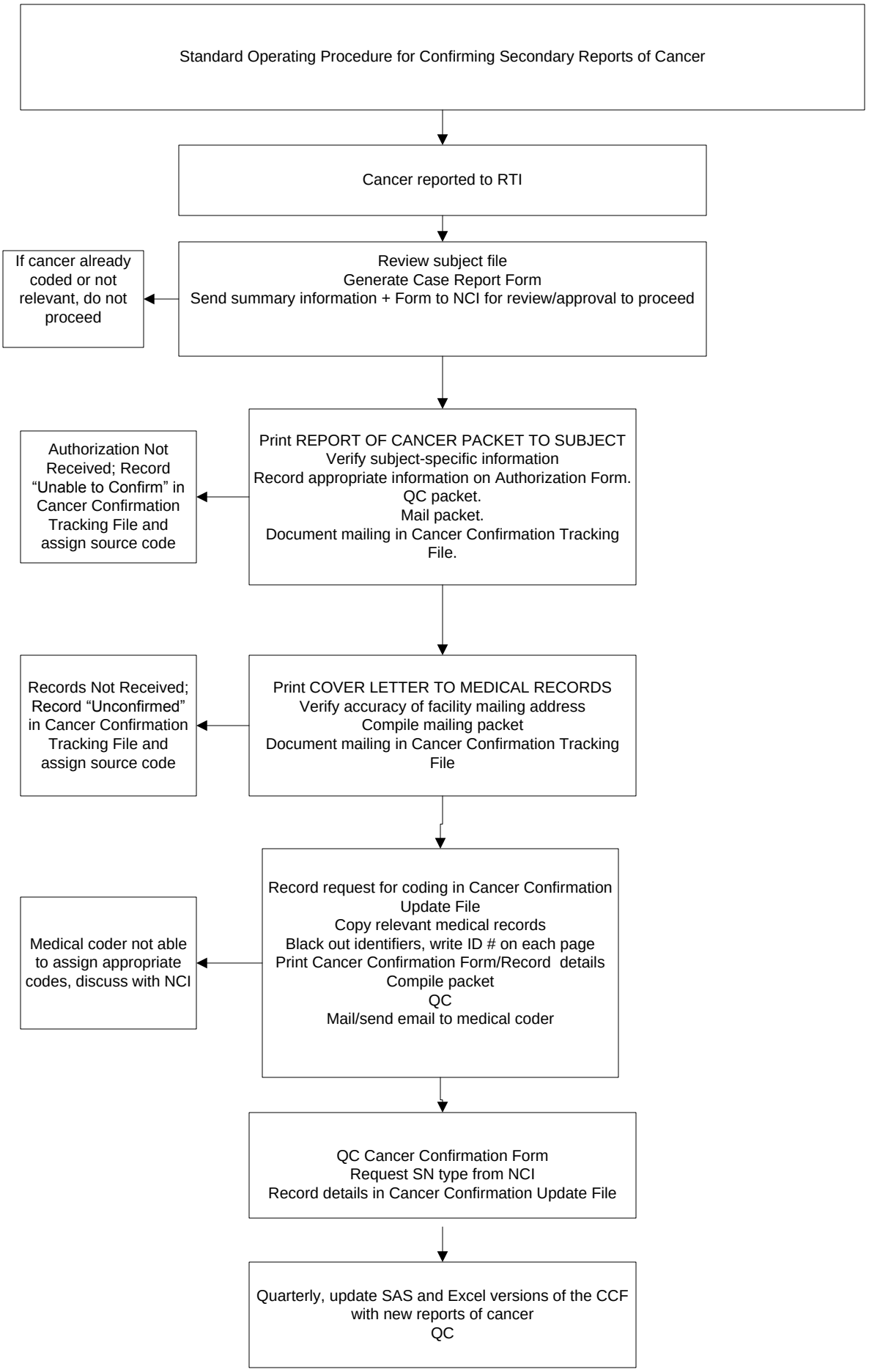
Name	**Relationship to you**	**Complete address/telephone number**

Thank you for completing this questionnaire. Your participation in this study has made a significant contribution to the study.

Please share any comments or concerns below.

Comments:

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____


Standard Operating Procedure for Confirming Secondary Reports of Cancer
Cancer reported to RTI
Review subject file
Generate Case Report Form
Send summary information + Form to NCI for review/approval to proceed
If cancer already coded or not relevant, do not proceed
Print REPORT OF CANCER PACKET TO SUBJECT
Verify subject-specific information
Record appropriate information on Authorization Form.
QC packet.
Mail packet.
Document mailing in Cancer Confirmation Tracking File.
Authorization Not Received; Record "Unable to Confirm" in Cancer Confirmation Tracking File and assign source code
Print COVER LETTER TO MEDICAL RECORDS
Verify accuracy of facility mailing address
Compile mailing packet
Document mailing in Cancer Confirmation Tracking File
Records Not Received; Record "Unconfirmed" in Cancer Confirmation Tracking File and assign source code
Record request for coding in Cancer Confirmation Update File
Copy relevant medical records
Black out identifiers, write ID # on each page
Print Cancer Confirmation Form/Record details
Compile packet
QC
Mail/send email to medical coder
Medical coder not able to assign appropriate codes, discuss with NCI
QC Cancer Confirmation Form
Request SN type from NCI
Record details in Cancer Confirmation Update File
Quarterly, update SAS and Excel versions of the CCF with new reports of cancer
QC

May 2013

NHL Subtype Paper Authors

Updated 12-18-2013

* please cc Lindsay (mortonli@mail.nih.gov) on emails for coordination purposes.

Acute lymphoblastic leukemia/lymphoma (ALL)			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Chris Skibola	cskibola@uab.edu	UCSF
Paper Leader	Dennis Weisenburger	dweisenburger@coh.org	*pathology*
Analyst	Susan Slager	Slager.Susan@mayo.edu	Mayo
Analyst	Beth Larabee	Larrabee.Beth@mayo.edu	
Authors	Sonja Berndt	berndts@mail.nih.gov	NCI-SEER
	Lindsay Morton	mortonli@mail.nih.gov	NCI-SEER
	Josh Sampson	sampsonjn@mail.nih.gov	
	Tracy Lightfoot	Tracy.Lightfoot@egu.york.ac.uk	UK

Burkitt lymphoma/leukemia (BL)			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Sam Mbulaiteye	mbulaits@mail.nih.gov	NCI-SEER
Paper Leader	Dennis Weisenburger	dweisenburger@coh.org	*pathology*
Analyst	Lindsay Morton	mortonli@mail.nih.gov	NCI-SEER
Analyst	Josh Sampson	sampsonjn@mail.nih.gov	
Authors	Ellen Chang	ellen@post.harvard.edu	SCALE
	Rob Newton	rob.newton@egu.york.ac.uk	UK
	Jennifer Kelly	Jennifer_kelly@urmc.rochester.edu	University of Rochester
	Jonathan Friedberg	jonathan_friedberg@urmc.rochester.edu	University of Rochester *(clinical)*
	Wendy Cozen	wcozen@usc.edu	NCI-SEER
	Rafael Marcos-Gragera	rmarcos@iconcologia.net	*pathology*
	Laura Costas Caudet	lcostas@iconcologia.net	EpiLymph
	Susan Slager	Slager.Susan@mayo.edu	Mayo
	Brenda Birmann	nhbmb@channing.harvard.edu	
	Silvia de Sanjosé	s.sanjose@iconcologia.net	EpiLymph

Chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL)			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Silvia de Sanjosé	s.sanjose@iconcologia.net	EpiLymph
Paper Leader	Susan Slager	Slager.Susan@mayo.edu	Mayo Clinic
Analyst	Beth Larabee	Larrabee.Beth@mayo.edu	
Authors	Jacqueline Clavel	jacqueline.clavel@inserm.fr	Engela
	Alain Monnereau	monnereau@bergonie.org	Engela
	Yolanda Benavente Moreno	ybenavente@iconcologia.net	EpiLymph

	Delphine Casabonne	dcasabonne@iconcologia.net	EpiLymph
	Pierluigi Cocco	coccop@medicina.unica.it	EpiLymph
	Marc Maynadié	mmaynadie@chu-dijon.fr	EpiLymph
	Alexandra Nieters	alexandra.nieters@uniklinik-freiburg.de	EpiLymph
	Adele Seniori Costantini	a.seniori@ispo.toscana.it	Italy
	Tracy Lightfoot	Tracy.Lightfoot@egu.york.ac.uk	UK
	Qing Lan	qingl@mail.nih.gov	Connecticut/ Yale University
	Aaron Blair	blaira@mail.nih.gov	Iowa/Minnesota; Kansas; NCI-SEER; Nebraska, older
	Sheila Hoar Zahm	zahms@mail.nih.gov	Iowa/Minnesota; Kansas; NCI-SEER; Nebraska, older
	Tim Call	call.timothy@mayo.edu	Mayo Clinic *(clinical)*
	Tina Clarke Dur	tina@nccc.org	*pathology*
	Lindsay Morton	mortonli@mail.nih.gov	NCI-SEER
	Josh Sampson	sampsonjn@mail.nih.gov	
	Roel Vermeulen	R.C.H.Vermeulen@uu.nl	

Diffuse large B-cell lymphoma (DLBCL)			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Jim Cerhan	cerhan.james@mayo.edu	Mayo Clinic
Paper Leader	Chris Skibola	cskibola@uab.edu	UCSF
Analyst	Lindsay Morton	mortonli@mail.nih.gov	
Analyst	Josh Sampson	sampsonjn@mail.nih.gov	
Authors	Anne Kricker	annek@health.usyd.edu.au	New South Wales
	Jacqueline Clavel	jacqueline.clavel@inserm.fr	Engela
	Alain Monnereau	monnereau@bergonie.org	Engela
	Alexandra Nieters	alexandra.nieters@uniklinik-freiburg.de	EpiLymph
	Lucia Miligi	l.miligi@ispo.toscana.it	Italy
	Luigino Dal Maso	dalmaso@cro.it	Northern Italy
	Tracy Lightfoot	Tracy.Lightfoot@egu.york.ac.uk	UK
	Nathaniel Rothman	rothmann@mail.nih.gov	NCI-SEER
	Sophia Wang	sowang@coh.org	NCI-SEER
	Aaron Blair	blaira@mail.nih.gov	Iowa/Minnesota; Kansas; NCI-SEER; Nebraska, older
	Chris Flowers	crflowe@emory.edu	*clinical*
	James Foran	foran.james@mayo.edu	*clinical*
	Ora Paltiel	ora@vms.huji.ac.il	*clinical*
	Susan Slager	Slager.Susan@mayo.edu	Mayo

Follicular lymphoma (FL)

Role	Name	Email Address	Study
Paper Leader	Brian Chiu	bchiu@uchicago.edu	Nebraska, newer
Paper Leader	Martha Linet	linetm@mail.nih.gov	NCI-SEER
Analyst	Silvia de Sanjosé	s.sanjose@iconcologia.net	
Analyst	Yolanda Benavente Moreno	ybenavente@iconcologia.net	
Authors	Claire Vajdic	claire.vajdic@unsw.edu.au	New South Wales
	Jacqueline Clavel	jacqueline.clavel@inserm.fr	Engela
	Alain Monnereau	monnereau@bergonie.org	Engela
	Pierluigi Cocco	coccop@medicina.unica.it	EpiLymph
	Paolo Boffetta	paolo.boffetta@mssm.edu	EpiLymph
	Lucia Miligi	l.miligi@ispo.toscana.it	Italy
	Alex Smith	alex.smith@egu.york.ac.uk	UK
	Tongzhang Zheng	tongzhang.zheng@yale.edu	Connecticut/Yale University
	Jim Cerhan	cerhan.james@mayo.edu	Mayo Clinic
	Anneclaire De Roos	aderoos@drexel.edu	NCI-SEER
	Chris Skibola	cskibola@uab.edu	UCSF
	Jennifer Kelly	Jennifer_kelly@urmc.rochester.edu	University of Rochester
	Aaron Blair	blaira@mail.nih.gov	Iowa/Minnesota; Kansas; NCI-SEER; Nebraska, older
	Tina Clarke Dur	tina@nccc.org	*pathology*
	Chris Flowers	crflowe@emory.edu	*clinical*
	Dennis Weisenburger	dweisenburger@coh.org	*pathology*
	Susan Slager	Slager.Susan@mayo.edu	Mayo
	Lindsay Morton	mortonli@mail.nih.gov	NCI-SEER
	Josh Sampson	sampsonjn@mail.nih.gov	

Hairy cell leukemia (HCL)			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Jacqueline Clavel	jacqueline.clavel@inserm.fr	Engela
Paper Leader	Alain Monnereau	monnereau@bergonie.org	Engela
Analyst	Susan Slager	Slager.Susan@mayo.edu	
Analyst	Beth Larabee	Larrabee.Beth@mayo.edu	
Authors	Ann Maree Hughes	AnnMaree.Hughes@anu.edu.au	New South Wales
	Anthony Staines	anthony.staines@dcu.ie	EpiLymph
	Bengt Glimelius	Bengt.glimelius@onkologi.uu.se	SCALE *(clinical)*
	Alex Smith	alex.smith@egu.york.ac.uk	UK
	Tom Habermann	habermann.thomas@mayo.edu	Mayo Clinic (*clinical)*
	Sonja Berndt	berndts@mail.nih.gov	NCI-SEER
	Pr Xavier Troussard	xavier.troussard@chu-caen.fr	Engela *(pathology)*
	Lindsay Morton	mortonli@mail.nih.gov	
	Josh Sampson	sampsonjn@mail.nih.gov	

Lymphoplasmacytic lymphoma/ Waldenström macroglobulinemia (LPL/WM)			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Claire Vajdic	claire.vajdic@unsw.edu.au	New South Wales
Paper Leader	Martha Linet	linetm@mail.nih.gov	NCI-SEER
Analyst	Susan Slager	Slager.Susan@mayo.edu	Mayo
Analyst	Beth Larabee	Larrabee.Beth@mayo.edu	
Authors	Anthony Staines	anthony.staines@dcu.ie	EpiLymph
	Alex Smith	alex.smith@egu.york.ac.uk	UK
	Angela Brooks-Wilson	abrooks-wilson@bcgsc.ca	British Columbia
	Steve Ansell	ansell.stephen@mayo.edu	Mayo Clinic *(clinical)*
	Ola Landgren	landgreo@mail.nih.gov	NCI-SEER *(clinical)*
	Mary Lou McMaster	mcmastem@mail.nih.gov	NCI-SEER
	Ahmet Dogan	Dogan.Ahmet@mayo.edu	Mayo Clinic *(pathology)*
	Lindsay Morton	mortonli@mail.nih.gov	NCI-SEER
	Josh Sampson	sampsonjn@mail.nih.gov	

Mantle cell lymphoma (MCL)			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Karin Ekström Smedby	Karin.Ekstrom.Smedby@ki.se	SCALE
Paper Leader	Lindsay Morton	mortonli@mail.nih.gov	NCI-SEER
Analyst	Susan Slager	Slager.Susan@mayo.edu	Mayo
Analyst	Beth Larabee	Larrabee.Beth@mayo.edu	
Authors	Jenny Turner	jtur8838@bigpond.net.au	New South Wales *(pathology)*
	Marc Maynadié	mmaynadie@chu-dijon.fr	EpiLymph *(hematopathologist)*
	Henrik Hjalgrim	hhj@ssi.dk	SCALE
	Eve Roman	Eve.Roman@egu.york.ac.uk	UK
	Tom Habermann	habermann.thomas@mayo.edu	Mayo Clinic *(clinical)*
	Sonja Berndt	berndts@mail.nih.gov	NCI-SEER
	Paige Bracci	paige.bracci@ucsf.edu	UCSF, older
	Chris Flowers	crflowe@emory.edu	*clinical*
	Dennis Weisenburger	dweisenburger@coh.org	*pathology*
	Josh Sampson	sampsonjn@mail.nih.gov	

Marginal zone lymphoma (MZL)			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Paige Bracci	paige.bracci@ucsf.edu	UCSF
Paper Leader	Silvia de Sanjosé	s.sanjose@iconcologia.net	
Analyst	Yolanda Benavente Moreno	ybenavente@iconcologia.net	
Authors	Jenny Turner	jtur8838@bigpond.net.au	New South Wales
	Marc Maynadié	mmaynadie@chu-dijon.fr	EpiLymph *(hematopathologist)*
	Bengt Glimelius	Bengt.glimelius@onkologi.uu.se	SCALE *(clinical)*
	Eve Roman	Eve.Roman@egu.york.ac.uk	UK
	John Spinelli	jspinelli@bccrc.ca	British Columbia
	Ahmet Dogan	Dogan.Ahmet@mayo.edu	Mayo Clinic *(pathology)*
	Herve Ghesquieres	Ghesquieres.Herve@mayo.edu; herve.ghesquieres@lyon.unicancer.fr	Mayo Clinic *(clinical)*
	Sophia Wang	sowang@coh.org	NCI-SEER
	Ora Paltiel	ora@vms.huji.ac.il	*clinical*
	Josh Sampson	sampsonjn@mail.nih.gov	
	Lindsay Morton	mortonli@mail.nih.gov	NCI-SEER
	Susan Slager	Slager.Susan@mayo.edu	Mayo

Mycosis fungoides/Sézary syndrome (MF/SS)			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Yawei Zhang	yawei.zhang@yale.edu	Connecticut/ Yale University
Paper Leader	Briseis Aschebrook-Kilfoy	brisa@uchicago.edu	Nebraska, newer
Analyst	Lindsay Morton	mortonli@mail.nih.gov	NCI-SEER
Analyst	Josh Sampson	sampsonjn@mail.nih.gov	
Authors	Claire Vajdic	claire.vajdic@unsw.edu.au	New South Wales
	Marshall Kadin	mkadin@chartercare.org	*pathology*
	Pierluigi Cocco	coccop@medicina.unica.it	EpiLymph
	Carlo La Vecchia	carlo.lavecchia@marionegri.it	Northern Italy
	Ellen Chang	ellen@post.harvard.edu	SCALE
	Rob Newton	rob.newton@egu.york.ac.uk	UK
	John Spinelli	jspinelli@bccrc.ca	British Columbia
	Andy Feldman	feldman.andrew@mayo.edu	Mayo Clinic *(pathology)*
	Carol Kasten	kastenca@mail.nih.gov	
	Susan Slager	Slager.Susan@mayo.edu	
	Sophia Wang	sowang@coh.org	NCI-SEER

Peripheral T-cell lymphoma (PTCL)			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Sophia Wang	sowang@coh.org	NCI-SEER
Paper Leader	Dennis Weisenburger	dweisenburger@coh.org	*pathology*
Analyst	Lindsay Morton	mortonli@mail.nih.gov	NCI-SEER
Analyst	Josh Sampson	sampsonjn@mail.nih.gov	
Authors	Ann Maree Hughes	AnnMaree.Hughes@anu.edu.au	New South Wales
	Paolo Boffetta	paolo.boffetta@mssm.edu	EpiLymph
	Ellen Chang	ellen@post.harvard.edu	SCALE
	Mads Melbye	mme@ssi.dk	SCALE
	Rob Newton	rob.newton@egu.york.ac.uk	UK
	Qing Lan	qingl@mail.nih.gov	Connecticut/ Yale University
	Yawei Zhang	yawei.zhang@yale.edu	Connecticut/ Yale University
	Steve Ansell	ansell.stephen@mayo.edu	Mayo Clinic *(clinical)*
	Andy Feldman	feldman.andrew@mayo.edu	Mayo Clinic *(pathology)*
	Chris Flowers	crflowe@emory.edu	*clinical*
	Marshall Kadin	mkadin@chartercare.org	*pathology*
	Susan Slager	Slager.Susan@mayo.edu	

Paper comparing all subtypes			
Role	**Name**	**Email Address**	**Study**
Paper Leader	Lindsay Morton	mortonli@mail.nih.gov	NCI-SEER
Paper Leader	Josh Sampson	sampsonjn@mail.nih.gov	
Paper Leader	Susan Slager	Slager.Susan@mayo.edu	Mayo Clinic
Paper Leader	Martha Linet	linetm@mail.nih.gov	NCI-SEER
Paper Leader	Dennis Weisenburger	dweisenburger@coh.org	*pathology*
Paper Leader	Jim Cerhan	cerhan.james@mayo.edu	Mayo Clinic
Authors	Bruce Armstrong	brucea@health.usyd.edu.au	New South Wales
	Jenny Turner	jtur8838@bigpond.net.au	New South Wales *(pathology)*
	John Kaldor	jkaldor@nchecr.unsw.edu.au	New South Wales
	Jacqueline Clavel	jacqueline.clavel@inserm.fr	Engela
	Alain Monnereau	monnereau@bergonie.org	Engela
	Paolo Boffetta	paolo.boffetta@mssm.edu	EpiLymph
	Marc Maynadié	mmaynadie@chu-dijon.fr	EpiLymph
	Anthony Staines	anthony.staines@dcu.ie	EpiLymph
	Carlo La Vecchia	carlo.lavecchia@marionegri.it	Northern Italy
	Eve Roman	Eve.Roman@egu.york.ac.uk	UK
	John Spinelli	jspinelli@bccrc.ca	British Columbia
	Anneclaire De Roos	aderoos@drexel.edu	NCI-SEER

	Patricia Hartge	hartgep@mail.nih.gov	NCI-SEER
	Nathaniel Rothman	rothmann@mail.nih.gov	NCI-SEER
	Sophia Wang	sowang@coh.org	NCI-SEER
	Brian Chiu	bchiu@uchicago.edu	Nebraska, newer
	Brenda Birmann	nhbmb@channing.harvard.edu	

Division of Cancer Epidemiology and Genetics

Radiation Epidemiology Branch

9609 Medical Center Drive,
Room 7E-454, MSC 9778
Bethesda, MD 20892-9778
email: mortonli@mail.nih.gov
phone:+1-240-276-7377
fax: +1-240-276-7840

U.S. Department of
Health and Human Services
National Institutes of Health

February 17, 2014

Carmen J. Allegra, M.D.
Editor-in-Chief
Journal of the National Cancer Institute Monographs
Oxford University Press, Suite 500
8120 Woodmont Avenue
Bethesda, MD 20814-2743

Dear Dr. Allegra,

The International Lymphoma Epidemiology Consortium (InterLymph) proposes publication of 13 manuscripts describing the risk factors for non-Hodgkin lymphoma (NHL) as a *Journal of the National Cancer Institute Monograph*. Exemplifying the advances in cancer epidemiology that can be achieved by consortial collaboration, we pooled data from 20 case-control studies from North America, Europe, and Australia, including a total of 17,471 NHL cases and 23,096 controls, to comprehensively evaluate medical history, lifestyle, family history, and occupational risk factors across histologic subtypes of NHL. These analyses represent the first large-scale investigations of the etiology of a number of rare NHL subtypes defined in the modern World Health Organization era. Additionally, for more common subtypes, we provide the first analyses to distinguish the etiologic contributions across multiple risk factors. Finally, we quantitatively compare risk factors across subtypes to understand the heterogeneity in the etiology of NHL subtypes. Our results represent a unique and fundamental contribution to our understanding of lymphomagenesis that should be of high impact and great interest to the broad clinical and scientific audience of the *Journal of the National Cancer Institute*.

We have included the following materials to further describe the project and the proposed contents of the monograph:

- Background, project rationale, and methods overview (Pages 2-4)
- Characteristics of the studies and participants included in the project (Pages 5-9)
- List of the 13 proposed manuscripts and their lead authors (Pages 10-11)
- Draft abstracts from 4 manuscripts for selected NHL subtypes (Pages 12-15)
- Draft abstract, tables, and figures from the final paper quantifying etiologic heterogeneity across NHL subtypes (Pages 16-26)

We would appreciate the opportunity to discuss the possibility of publishing the InterLymph NHL Subtypes Project as a *Journal of the National Cancer Institute Monograph*, and we look forward to hearing from you.

Sincerely yours,

Lindsay M. Morton, Ph.D.
National Cancer Institute

Susan L. Slager, Ph.D.
Mayo Clinic

BACKGROUND AND RATIONALE

Non-Hodgkin lymphoma (NHL), which affects over 86,000 individuals in the United States annually [1], is actually composed of numerous diseases with distinctive morphologic, immunophenotypic, genetic, and clinical features. The strongest known NHL risk factor is severe immunodeficiency, but the etiologies of most lymphomas remain unknown.

In 2001, the World Health Organization introduced an international, consensus-based classification for hematologic malignancies, providing the first reliable framework for consistently defining the various subtypes of NHL [2]. Since that time, descriptive and analytic epidemiologic studies have increasingly pointed to both heterogeneity and commonality in the etiology of lymphoma subtypes. Population registry data reveal striking differences in incidence among lymphoma subtypes by age, sex, race, and calendar year [3-5]. Analytic studies show that certain infections are associated with specific NHL subtypes [e.g., human T-cell lymphotropic virus, type I (HTLV-I) with adult T-cell leukemia/lymphoma [6], and *Helicobacter pylori* with gastric mucosa-associated lymphoid tissue (MALT) NHL [7]], whereas other infections are associated with multiple lymphoma subtypes [e.g., human immunodeficiency virus (HIV) [8] and hepatitis C virus (HCV) [9]]. Variation in the magnitude of risks across NHL subtypes also is clearly evident for patients with a history of autoimmune conditions [10], family history of cancer [11], and certain common genetic variants [12, 13]. In contrast, sun exposure and alcohol consumption appear to affect risk for all NHL subtypes [14, 15].

Despite the advances in our understanding of the etiology of some NHL subtypes in the last decade, little is known about the etiology of rare NHL subtypes, and never before have multiple key risk factors been modeled simultaneously for any NHL subtype. Because individual epidemiologic studies do not have sufficient sample size to address these scientific questions, and because new studies have joined the consortium since the publication of most previous pooled analyses of specific risk factors [9-18], InterLymph undertook a consortium-wide initiative, the "InterLymph NHL Subtypes Project," with the following primary aims:

1. To comprehensively evaluate risk factor profiles across NHL subtypes in a multivariate setting, and
2. To quantitatively assess etiologic heterogeneity across NHL subtypes.

METHODS OVERVIEW

The InterLymph NHL Subtypes Project pooled individual-level data from 20 case-control studies across North America, Europe, and Australia, including a total of 17,471 NHL cases and 23,096 controls (Table 1). Nearly 75% of the participants are derived from the 14 population-based studies, with the remaining participants from hospital-based studies. The pooled study population is 58% male and 93% white (Table 2).

For 69% of the cases, the NHL subtype was classified using the World Health Organization classification or similar approach (Table 3). Within each study, centralized pathology review to histologically confirm incident NHL diagnoses was conducted by at least one expert hematopathologist in all studies except NCI-SEER and Italy multi-center, for which cases were reviewed by local diagnosing pathologists.

All NHL histologic subtypes with at least 100 cases are included in the InterLymph NHL Subtypes Project. The most commonly diagnosed NHL subtypes were diffuse large B-cell lymphoma (N=4667), follicular lymphoma (N=3530), and chronic lymphocytic leukemia/small lymphocytic lymphoma (N=2440) (Table 4). Other subtypes included in the project are marginal zone lymphoma (N=1052), peripheral T-cell lymphoma (N=584), mantle cell lymphoma (N=557), lymphoplasmacytic lymphoma/Waldenström macroglobulinemia (N=374), and mycosis fungoides/Sézary syndrome (N=324), as well as the rare NHL subtypes Burkitt lymphoma/leukemia (N=295), hairy cell leukemia (N=154), and acute lymphoblastic leukemia/lymphoma (N=152).

In each of the 20 contributing studies, data on putative risk factors for NHL were ascertained by in-person or telephone interviews (typically computer-assisted) and self-reported questionnaires. Risk factors selected for inclusion in this analysis were the available medical history, lifestyle, family history, and occupational risk factors from at least four studies (Table 5). Individual-level data were sent from each study for each exposure category, and analytic variables for the pooled analysis were harmonized centrally at the InterLymph Data Coordinating Center at the Mayo Clinic under the leadership of Dr. Susan Slager. Data were then transferred to the Division of Cancer Epidemiology and Genetics, National Cancer Institute, for centralized data analysis under the leadership of Drs. Lindsay Morton and Joshua Sampson. Analyses include multivariate modeling of key risk factors for each NHL subtype, as well as cutting-edge statistical methodology for comparing etiologic

heterogeneity among NHL subtypes. Throughout the project, data harmonization and analyses have been guided by an interdisciplinary group of epidemiologists, pathologists, biostatisticians, and clinicians, with representation from each contributing study as well as each of the InterLymph Working Groups (Pathology, Lifestyle & Environment, and Immunity & Infection).

We propose to include 13 manuscripts, beginning with an overview of the project (including rationale and methods), 11 subtype-specific papers, and a final paper that quantitatively compares risk factors across NHL subtypes. The names and affiliations of the leaders for each manuscript are listed on pages 10-11. All manuscripts have been drafted and are currently undergoing rigorous internal review within the InterLymph Consortium.

Table 1. Characteristics of studies included in the InterLymph NHL Subtypes Project

Region Study (reference)	Location	Years of diagnosis	Design	Participation % Controls	% Cases
North America					
British Columbia [19]	Vancouver, Victoria, British Columbia (Canada)	2000-2004	Population-based	46	79
Iowa/Minnesota [20][†]	Iowa, Minnesota (USA)	1981-1983	Population-based	81	87
Kansas [21][‡]	Kansas (USA)	1976-1982	Population-based	94	96
Los Angeles [22][‡]	Los Angeles, California (USA)	1989-1992	Population-based	69	45
Mayo Clinic [23, 24][†]	Iowa, Minnesota, Wisconsin (USA)	2002-2008	Clinic-based	69	69
NCI-SEER [25]	Detroit, Michigan; Iowa; Los Angeles, California; Seattle, Washington (USA)	1998-2000	Population-based	52	76
Nebraska (older) [26][†]	Nebraska (USA)	1983-1986	Population-based	87	91
Nebraska (newer) [27]	Nebraska (USA)	1999-2002	Population-based	77	73
UCSF1 [28]	San Francisco, California (USA)	1988-1995	Population-based	78	72
UCSF2*	San Francisco, California (USA)	2001-2006	Population-based	68	70
Univ. of Rochester [29]	Rochester, New York (USA)	2005-2007	Hospital-based	78	96
Yale [30]	Connecticut (USA)	1995-2001	Population-based	47-69	72
Europe					
Engela [31][†,‡]	Bordeaux, Brest, Caen, Lille, Nantes, Toulouse (France)	2000-2004	Hospital-based	93	97
EpiLymph [32][†,‡]	Spain; France; Germany; Italy; Ireland; Czech Republic	1998-2004	Population-based (Italy, Germany), otherwise hospital-based	44-96	82-93
Italy multi-center [33][†]	Firenze, Forli, Imperia, Latina, Novara, Ragusa, Siena, Torino, Varese, Vercelli, Verona (Italy)	1990-1993	Population-based	74	82
Italy (Aviano-Milan) [34]	Aviano, Milan (Italy)	1983-1992	Hospital-based	>97	>97
Italy (Aviano-Naples) [35]	Aviano, Naples (Italy)	1999-2002	Hospital-based	91	97
SCALE [36][†]	Denmark; Sweden	1999-2002	Population-based	71	81
United Kingdom [37][†,‡]	Lancaster/South Lakeland, South England, Yorkshire (United Kingdom)	1998-2001	Population-based	71	75
Australia					
New South Wales [38]	New South Wales, Australian Capital Territory (Australia)	2000-2001	Population-based	61	85

Abbreviations: National Cancer Institute-Surveillance, Epidemiology, and End Results Program (NCI-SEER); Scandinavian Lymphoma Etiology Study (SCALE); University of California, San Francisco (UCSF).

* This recently completed study has not yet published results but had similar methodology to a previous study conducted in the same region [28].

[†] Studies were designed to investigate other malignancies in addition to NHL: Engela: Hodgkin lymphoma, multiple myeloma, and lymphoproliferative syndromes; EpiLymph: Hodgkin lymphoma and multiple myeloma; Iowa/Minnesota: leukemia; Italy: Hodgkin lymphoma, multiple myeloma, and soft-tissue sarcoma; Mayo: Hodgkin lymphoma; Nebraska (older): Hodgkin lymphoma and multiple myeloma; SCALE: Hodgkin lymphoma; United Kingdom: Hodgkin lymphoma.

[‡] Studies individually matched controls to cases. The remaining studies frequency matched controls to cases by age (typically in 5-year groups), sex, and race.

Table 2. Selected characteristics of controls from participating studies in the InterLymph NHL Subtypes Project

Region	Controls		Race/Ethnicity					Age (years) †
Study	N	%Male *	%White	%Black	%Asian	%Hispanic	%Other ‡	Median (min-max)
North America								
British Columbia	845	53	77	0	17	0	7	60(20-80)
Iowa/Minnesota	1245	100	100	0	0	0	0	68(30-97)
Kansas	948	100	96	0	0	1	3	63(18-98)
Los Angeles	375	49	75	5	3	16	0	54(17-79)
Mayo Clinic	1314	54	98	0	0	1	1	63(21-94)
NCI-SEER	1055	52	78	14	2	5	1	61(20-74)
Nebraska (older)	1432	51	99	0	0	0	1	68(19-98)
Nebraska (newer)	533	53	96	2	1	1	1	59(20-76)
UCSF1	2402	65	81	6	5	7	2	53(21-74)
UCSF2	457	59	94	0	0	6	0	59(20-84)
Univ. of Rochester	139	44	88	6	1	3	1	52(21-85)
Yale	717	0	92	3	0	4	1	64(23-86)
Europe								
Engela	722	62	100	0	0	0	0	55(18-76)
EpiLymph	2460	54	97	0	0	0	3	59(17-76)
Italy multi-center	1771	52	100	0	0	0	0	58(19-75)
Italy (Aviano-Milan)	1157	61	100	0	0	0	0	57(17-85)
Italy (Aviano-Naples)	504	68	100	0	0	0	0	63(18-83)
SCALE	3187	55	95	0	0	0	5	59(18-76)
United Kingdom	1139	55	100	0	0	0	0	53(16-69)
Australia								
New South Wales	694	57	87	0	2	0	11	57(21-74)

Abbreviations: National Cancer Institute-Surveillance, Epidemiology, and End Results Program (NCI-SEER); Scandinavian Lymphoma Etiology Study (SCALE); University of California, San Francisco (UCSF).

* The Yale study was restricted to women, whereas the Kansas and Iowa/Minnesota studies were restricted to men.

† Age at diagnosis for cases and interview for controls. All studies excluded children (with varying minimum ages), and most studies excluded the elderly (with varying maximum ages), except for the Iowa/Minnesota, Kansas, Mayo Clinic, Nebraska (older), and University of Rochester studies, which did not have an upper age limit.

‡ Includes individuals with other specified races as well as unknown race. For studies with predominantly white, non-Hispanic populations, individuals with unknown race were assumed to be white [Iowa/Minnesota, Kansas, Mayo Clinic, NCI-SEER (Seattle, Iowa study centers), Nebraska (older), Nebraska (newer), UCSF2, Univ. of Rochester, Yale, Engela, SCALE, United Kingdom].

Table 3. NHL subtype classification and pathology review

Region Study	NHL Subtype Classification	Pathology review *		
		Pathology Reports	Diagnostic Slides	Additional Immunophenotyping
North America				
British Columbia	WHO/ICD-O-3	✓	✓	As needed
Iowa/Minnesota	WF	✓	✓	-
Kansas	WF	✓	✓	-
Los Angeles	WF	✓	✓	-
Mayo Clinic	WHO	✓	✓	-
NCI-SEER	WHO/ICD-O-2	-	-	-
Nebraska (older)	WF	✓	✓	-
Nebraska (newer)	WHO	✓	✓	-
UCSF1	WF	✓	✓	-
UCSF2	WHO/ICD-O-3	✓	✓	-
Univ. of Rochester	WHO	✓	✓	-
Yale	REAL	✓	✓	-
Europe				
Engela	WHO/ICD-O-3	✓	-	-
EpiLymph	WHO	20% of cases	20% of cases	As needed
Italy multi-center	WF	-	-	-
Italy (Aviano-Milan)	WF	✓	✓	-
Italy (Aviano-Naples)	WHO	✓	-	-
SCALE	WHO	✓	✓	-
United Kingdom	WHO/ICD-O-3	✓	-	-
Australia				
New South Wales	WHO/ICD-O-3	✓	44% of cases	As needed

Abbreviations: International Classification of Diseases for Oncology (ICD-O); National Cancer Institute-Surveillance, Epidemiology, and End Results Program (NCI-SEER); Revised European-American Lymphoma (REAL) classification; Scandinavian Lymphoma Etiology Study (SCALE); University of California, San Francisco (UCSF); Working Formulation (WF), World Health Organization (WHO).

* Centralized pathology review to histologically confirm NHL diagnoses was conducted by at least one expert hematopathologist in all studies except NCI-SEER and Italy multi-center, for which cases were diagnosed by local pathologists.

Table 4. Distribution of the InterLymph NHL Subtypes Project study population by case/control status, NHL subtype (for cases), and study

NHL Subtype Classification		**Cases**												
Study	**Controls**	**Total**	**DLBCL**	**FL**	**CLL/SLL**	**MZL**	**PTCL**	**MCL**	**LPL**	**MF/SS**	**BL**	**HCL**	**ALL**	**Other/NOS****
World Health Organization														
New South Wales	694	694	231	252	29 *	61	16	22	27	4	4	10	5	33
Mayo Clinic	1314	1120	210	271	376	68	32	56	22	9	7	0 †	1	68
British Columbia	845	833	218	228	42 *	101	33	50	42	42	10	0 †	6	61
Nebraska (newer)	533	387	103	123	29 *	35	12	16	5	7	12	0 †	1	44
United Kingdom	1139	828	326	236	0 †	141	50	40	0 †	15	0 †	0 †	0 †	20
NCI-SEER	1055	1316	413	318	133 *	106	55	50	28	26	12	0 †	0 †	175
EpiLymph	2460	1735	516	251	414	138	78	67	44	38	24	15	46	104
Yale	717	600	189	136	50 *	40	19	16	16	12	4	0 †	3	115
SCALE	3187	3055	796	586	752	117	121	148	116	41	31	63	15	269
Engela	722	567	174	101	132	20	18	25	21	0	11	36	8	21
Aviano-Naples	504	225	112	36	18 *	14	9	2	10	2	9	0 †	0 †	13
UCSF2	457	487	0 §	0 §	0 §	187	77	58	43	54	35	0 †	7	26
Univ. of Rochester ‖	139	129	32	45	7 *	24	8	7	0 †	2	0	0 †	0	4
Working Formulation														
Los Angeles ¶	375	376	151	41	0 ‡	0 ‡	1	0 ‡	0 †	0 ‡	50	0 †	16	117
UCSF1	2402	1302	509	352	0 ‡	0 ‡	0 ‡	0 ‡	0 †	47	28	0 †	12	354
Italy multi-center	1771	1911	407	159	214	0 ‡	55	0 ‡	0 †	25	23	30	12	986
Aviano-Milan	1157	429	47	55	0 ‡	0 ‡	0 ‡	0 ‡	0 †	0 ‡	9	0 †	10	308
Iowa/Minnesota	1245	866	112	195	244	0 ‡	0 ‡	0 ‡	0 ‡	0 ‡	18	0 †	6	291
Kansas	948	170	27	34	0 ‡	0 ‡	0 ‡	0 ‡	0 ‡	0 ‡	2	0 †	1	106
Nebraska (older)	1432	441	94	111	0 ‡	0 ‡	0 ‡	0 ‡	0 ‡	0 ‡	6	0 †	3	227
Total	23,096	17,471	4667	3530	2440	1052	584	557	374	324	295	154	152	3342

Abbreviations: acute lymphoblastic leukemia (ALL); Burkitt/Burkitt-like lymphoma/leukemia (BL); chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL); diffuse large B-cell lymphoma (DLBCL); follicular lymphoma (FL); hairy cell leukemia (HCL); lymphoplasmacytic lymphoma/Waldenström macroglobulinemia (LPL); mantle cell lymphoma (MCL); marginal zone lymphoma (MZL); mycosis fungoides/Sézary syndrome (MF/SS); National Cancer Institute-Surveillance, Epidemiology, and End Results Program (NCI-SEER); not otherwise specified (NOS); peripheral T-cell lymphoma (PTCL); Scandinavian Lymphoma Etiology Study (SCALE); University of California, San Francisco (UCSF).

* Included SLL but not CLL based on original study eligibility criteria.

† NHL subtype excluded based on original study eligibility criteria.

‡ NHL subtypes were classified according to the Working Formulation with varying levels of additional pathology data, and certain subtypes could not be distinguished reliably.

§ The three most common NHL subtypes were excluded from this recently completed study, pending publication of individual study results.

‖ Excluded cases with evident CNS involvement.

¶ Restricted to intermediate- or high-grade NHL, as defined by the Working Formulation.

** We grouped cases into NHL subtypes according to the World Health Organization classification [2, 39] using guidelines from the InterLymph Pathology Working Group [17, 40]. NHL subtypes with ≥100 cases in the pooled dataset were eligible for inclusion in this project. Other/NOS includes subtypes with <100 cases and poorly specified subtypes. The higher proportion of Other/NOS cases from the Working Formulation studies is due to low correspondence between certain Working Formulation and World Health Organization subtypes [17].

Table 5. Contribution of risk factor data for the InterLymph NHL Subtypes Project

Risk factor	British Columbia	Iowa/Minnesota	Kansas	Los Angeles	Mayo Clinic	NCI-SEER	Nebraska (newer)	Nebraska (older)	UCSF1	UCSF2	Univ. of Rochester	Yale	Engela	EpiLymph	Italy multi-center	Italy (Aviano-Milan)	Italy (Aviano-Naples)	SCALE	United Kingdom	New South Wales
Lifestyle risk factors																				
Anthropometric measures	✓		✓		✓	✓	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	
Alcohol consumption		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Cigarette smoking		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Physical activity	✓				✓	✓	✓				✓	✓		✓						
Personal hair dye use	✓					✓	✓	✓	✓			✓	✓	✓	✓					
Sun exposure	✓				✓	✓	✓				✓	✓	✓	✓				✓	✓	✓
Medical history risk factors																				
Autoimmune disease	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Atopy (allergy, asthma, hay fever, or eczema)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hepatitis C viral infection	✓								✓					✓			✓	✓		✓
History of blood transfusion	✓	✓		✓	✓	✓	✓		✓			✓	✓	✓			✓		✓	✓
Reproductive history	✓			✓	✓	✓	✓		✓			✓		✓		✓	✓		✓	
Oral contraceptive use	✓			✓	✓				✓			✓		✓		✓	✓			
Hormonal replacement therapy use	✓			✓	✓				✓			✓				✓	✓		✓	
Gastric/peptic ulcer disease	✓				✓	✓			✓	✓				✓		✓	✓	✓	✓	✓
Family history of hematologic malignancy	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	
History of farm exposure	✓	✓	✓		✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓		✓	✓
Occupational history	✓					✓	✓		✓			✓	✓	✓	✓				✓	✓

Abbreviations: National Cancer Institute-Surveillance, Epidemiology, and End Results Program (NCI-SEER); Scandinavian Lymphoma Etiology Study (SCALE); University of California, San Francisco (UCSF).

PROPOSED MANUSCRIPTS AND LEAD AUTHORS

1. Overview of the International Lymphoma Epidemiology Consortium (InterLymph) Non-Hodgkin Lymphoma Subtypes Project
   – **Lindsay Morton, Ph.D.**, Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD
   – **Susan Slager, Ph.D.**, Division of Biostatistics, Department of Health Sciences Research, College of Medicine, Mayo Clinic, Rochester, MN

2. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Diffuse Large B-Cell Lymphoma: The Interlymph NHL Subtypes Project
   – **James Cerhan, M.D., Ph.D.**, Division of Epidemiology, Department of Health Sciences Research, College of Medicine, Mayo Clinic, Rochester, MN
   – **Christine Skibola, Ph.D.**, Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA

3. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Follicular Lymphoma: The Interlymph NHL Subtypes Project
   – **Brian Chiu, Ph.D.**, Department of Health Studies, University of Chicago, Chicago, IL
   – **Martha Linet, M.D., M.P.H.**, Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD

4. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma: The Interlymph NHL Subtypes Project
   – **Silvia de Sanjosé, M.D.**, Institut Català d'Oncologia-Catalan Institute of Oncology, L'Hospitalet de Llobregat, Barcelona, Spain
   – **Susan Slager, Ph.D.**, Division of Biostatistics, Department of Health Sciences Research, College of Medicine, Mayo Clinic, Rochester, MN

5. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Marginal Zone Lymphoma: The Interlymph NHL Subtypes Project
   – **Paige Bracci, Ph.D., M.P.H., M.S.**, Department of Epidemiology and Biostatistics, School of Medicine, University of California, San Francisco, CA
   – **Silvia de Sanjosé, M.D.**, Institut Català d'Oncologia-Catalan Institute of Oncology, L'Hospitalet de Llobregat, Barcelona, Spain

6. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Peripheral T-Cell Lymphoma: The Interlymph NHL Subtypes Project
   – **Sophia Wang, Ph.D.**, Department of Cancer Etiology, City of Hope Beckman Research Institute, Duarte, CA
   – **Dennis Weisenburger, M.D.,** Department of Pathology, City of Hope National Medical Center, Duarte, CA

7. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Mantle Cell Lymphoma: The Interlymph NHL Subtypes Project
   - **Karin Ekström Smedby, Ph.D.**, Department of Medicine, Clinical Epidemiology Unit, Karolinska Institutet, Stockholm, Sweden
   - **Lindsay Morton, Ph.D.**, Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD

8. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Lymphoplasmacytic Lymphoma/Waldenström Macroglobulinemia: The Interlymph NHL Subtypes Project
   - **Claire Vajdic, Ph.D.**, Cancer Aetiology and Prevention Group, Prince of Wales Clinical School, University of New South Wales, Sydney, Australia
   - **Martha Linet, M.D., M.P.H.**, Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD

9. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Mycosis Fungoides and Sézary Syndrome: The Interlymph NHL Subtypes Project
   - **Yawei Zhang, M.D., Ph.D., M.P.H.**, Yale University School of Public Health, New Haven, CT
   - **Briseis Aschebrook-Kilfoy, Ph.D.**, University of Chicago, Chicago, IL

10. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Burkitt Lymphoma/Leukemia: The Interlymph NHL Subtypes Project
    - **Sam Mbulaiteye, M.D.**, Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD
    - **Dennis Weisenburger, M.D.,** Department of Pathology, City of Hope National Medical Center, Duarte, CA

11. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Hairy Cell Leukemia: The Interlymph NHL Subtypes Project
    - **Alain Monnereau, M.D., Ph.D.**, INSERM, CESP, University Paris South, Villejuif, France; Haematological Cancer Registry of Gironde, Bergonie Institute, Comprehensive Cancer Center of Bordeaux and South West, France
    - **Jacqueline Clavel**, **M.D., Ph.D.**, INSERM, CESP, University Paris South, Villejuif, France

12. Medical History, Lifestyle, Family History, and Occupational Risk Factors for Acute Lymphoblastic Leukemia/Lymphoma: The Interlymph NHL Subtypes Project
    - **Christine Skibola, Ph.D.**, Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA
    - **Dennis Weisenburger, M.D.,** Department of Pathology, City of Hope National Medical Center, Duarte, CA

13. Etiologic Heterogeneity Among NHL Subtypes: The Interlymph NHL Subtypes Project
    - **Lindsay Morton, Ph.D.**, Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD
    - **Joshua Sampson, Ph.D.**, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD

Medical History, Lifestyle, Family History, and Occupational Risk Factors for Diffuse Large B-Cell Lymphoma: The Interlymph NHL Subtypes Project

Background While risk factors for DLBCL have been suggested, their independent effects, modification by sex, and association with anatomic sites are largely unknown.

Methods In a pooled analysis of 4,667 cases and 22,639 controls from 19 studies, stepwise logistic regression identified the most parsimonious multivariate models for DLBCL overall and by sex. Risk factors by anatomic site were also assessed.

Results DLBCL was associated with B-cell activating autoimmune diseases (odds ratio(OR)=2.36, 95% confidence interval(CI) 1.80 to 3.09), HCV seropositivity (2.02, 1.47 to 2.76), family history of NHL (1.95, 1.54 to 2.47), higher young adult body mass index (BMI) (1.58, 1.12 to 2.23, for 35+ versus 18.5-22.4 kg/m^2), higher recreational sun exposure (0.78, 0.69 to 0.89), any atopic disorder (0.82, 0.76 to 0.89), and higher socioeconomic status (0.86, 0.79 to 0.94) in the final model. Additional risk factors for females were field crop/vegetable farm workers (1.78, 1.22 to 2.60), hairdressers (1.65, 1.12 to 2.41), seamstresses/embroiderers (1.49, 1.13 to 1.97), low adult BMI (0.46, 0.29 to 0.74, for <18.5 vs. 18.5-22.4 kg/m^2), hormone replacement therapy started age ≥50 years (0.68, 0.52 to 0.88), and oral contraceptive use before 1970 (0.78, 0.62 to 1.00); and for males were material handling equipment operators (1.58, 1.02 to 2.44), lifetime alcohol consumption (0.57, 0.44 to 0.75, for >400kg vs. nondrinker), and previous blood transfusion (0.69, 0.57 to 0.83). Autoimmune disease, atopy, and family history of NHL showed similar associations across selected anatomic sites, while smoking was associated with central nervous system, testicular and cutaneous DLBCLs; inflammatory bowel disease with gastrointestinal DLBCL; and farming and hair dye with mediastinal DLBCL.

Conclusions Our results support a complex and multifactorial etiology for DLBCL with some variation by sex and anatomic site.

Medical History, Lifestyle, Family History, and Occupational Risk Factors for Peripheral T-Cell Lymphoma: The Interlymph NHL Subtypes Project

Background Accounting for 10-15% of all NHLs in Western populations, peripheral T-cell lymphomas (PTCL) are the most common T-cell lymphomas, but they still relatively rare and, thus, little is known about their etiology. Our aim was to identify etiologic risk factors for PTCL overall, and for specific PTCL subtypes, by analyzing data from 15 epidemiologic studies participating in the InterLymph Consortium.

Methods A pooled analysis of individual-level data for 584 histologically-confirmed PTCL cases and 15,912 controls from case-control studies conducted in Europe, North America, and Australia was undertaken. Study questionnaires were harmonized to permit evaluation of a broad range of potential risk factors. Odds ratios (OR) and 95% confidence intervals (CI) were calculated using logistic regression.

Results Risk factors that significantly increased overall PTCL risk ($P<0.05$) included: a family history of hematologic malignancies; T-cell activating autoimmune conditions; eczema; smoking; and the occupational groups of textile worker or electric fitter. A personal history of allergies, having ever lived or worked on a farm, and alcohol consumption decreased PTCL risk. Differences were also observed for some PTCL subtypes, including: an increased risk of primary cutaneous anaplastic large-cell lymphoma (ALCL) among those with hay fever; elevated unspecified cutaneous PTCL risk among persons with asthma; and elevated risk of enteropathy-type PTCL among those with celiac disease, a history of peptic ulcer or blood tranfusion.

Conclusions Our pooled analyses identified a number of new risk factors for PTCL. Our results support etiologic differences by PTCL subtypes but require further validation in independent series.

Medical History, Lifestyle, Family History, and Occupational Risk Factors for Mantle Cell Lymphoma: The Interlymph NHL Subtypes Project

Background The etiology of mantle cell lymphoma (MCL), a distinctive subtype accounting for 2-10% of all non-Hodgkin lymphoma (NHL), is not known.

Methods We investigated associations with self-reported lifestyle, medical history, family history, and occupational risk factors in a pooled analysis of 557 patients with MCL and 13,766 controls from 13 case-control studies in Europe, North America, and Australia. Odds ratios (ORs) and 95% confidence intervals (CIs) associated with each exposure were examined using multivariable logistic regression models.

Results The median age of the MCL patients was 62 years and 76% were men. Risk of MCL was inversely associated with history of hay fever (OR=0.63, 95% CI 0.48 to 0.82), and the association was independent of other atopic diseases and allergies. A hematological malignancy among first-degree relatives was associated with a 2-fold increased risk of MCL (1.99, 1.39 to 2.84), which was stronger in men (2.21, 1.44 to 3.38) than women (1.61, 0.82 to 3.19). A modestly increased risk of MCL was also observed in association with ever having lived on a farm (1.40, 1.03 to 1.90). Unlike some other NHL subtypes, MCL risk was not statistically significantly associated with autoimmune disorders, tobacco smoking, alcohol intake, body mass index, or ultraviolet radiation.

Conclusions The novel observations of a possible role for atopy/allergy and farm life in risk of MCL, together with confirmatory evidence of a familial link, suggest a multifactorial etiology of immune-related environmental exposures and genetic susceptibility. These findings provide guidance for future research in MCL etiology.

Medical History, Lifestyle, Family History, and Occupational Risk Factors for Hairy Cell Leukemia: The Interlymph NHL Subtypes Project

Background Little is known about the etiology of hairy cell leukemia (HCL), a rare B-cell lymphoproliferative disorder with marked male predominance. Our aim was to identify key risk factors for HCL.

Methods A pooled analysis of individual-level data for 154 histologically-confirmed HCL cases and 8834 controls from five case-control studies, conducted in Europe and Australia, was undertaken for males and females separately and for both sexes combined.

Results The usual pattern for age and sex in HCL were observed, with a median age of 55 years and sex ratio of 3.6 males to females. Cigarette smoking was inversely associated with HCL (OR=0.51, 95% CI 0.37 to 0.71) with dose-response relationships observed for duration, frequency and lifetime cigarette smoking statistically associated (P_{trend}<0.001 in men). In contrast, farming occupation was significantly positively associated with HCL (OR=2.34, 95% CI 1.36 to 4.01), with a dose-response relationship observed for duration (OR=1.82, 95% CI 0.85 to 3.88 for <10 years and OR=2.98, 95% CI 1.50 to 5.93 for ≥10 years vs. never; P_{trend}=0.025). Adult height tended to be higher in cases compared to controls (OR=2.69, 95% CI 1.39 to 5.29 for highest vs. lowest sex-specific quartile). In the multivariate analysis, living or working on a farm, height and cumulative cigarette smoking were independently associated with HCL risk.

Conclusions Our observations of an increased risk of HCL from farming exposures and decrease risk from smoking exposures, independently of each other, support a multifactorial origin and an etiological specificity of HCL. The positive association with height needs replication.

Etiologic Heterogeneity Among NHL Subtypes: The Interlymph NHL Subtypes Project

Background Non-Hodgkin lymphoma (NHL) is composed of many biologically and clinically heterogeneous subtypes, but large-scale investigation of risk factor patterns across NHL subtypes is lacking.

Methods We pooled individual-level data from 17,471 NHL cases and 23,096 controls in 20 case-control studies from the International Lymphoma Epidemiology Consortium (InterLymph). Family history, medical history, lifestyle, and occupational risk factors were evaluated across NHL subtypes using a subset-based statistical approach and diagrammed in a tree, with nodes defined by a permutation-based minimum P value from all possible pairwise comparisons.

Results The greatest difference in risk factor patterns occurred between T-cell and B-cell lymphomas ($P_{NODE}<0.001$), with increased risks generally restricted to T-cell lymphomas for eczema, T-cell activating autoimmune diseases, family history of multiple myeloma, and occupation as a painter. Substantial heterogeneity also was observed among B-cell lymphomas ($P_{NODE}<0.001$). A cluster of risk factors including B-cell activating autoimmune diseases, hepatitis C virus seropositivity, family history of Hodgkin lymphoma, alcohol consumption, and occupation as a teacher was most strongly associated with marginal zone lymphoma, Burkitt/Burkitt-like lymphoma/leukemia, diffuse large B-cell lymphoma, and lymphoplasmacytic lymphoma/Waldenström macroglobulinemia. Heterogeneity across NHL subtypes also was observed for cigarette smoking, anthropometric measures, family history of leukemia, and history of blood transfusion. In contrast, modest and generally homogeneous risks across NHL subtypes were observed for family history of NHL, recreational sun exposure, hay fever, and allergy.

Conclusions We found significant heterogeneity of effect among the NHL subtypes for the risk factors considered herein, with only a few risk factors showing commonalities. Future research is needed to improve our understanding of the mechanisms of lymphomagenesis, investigate other risk factors (e.g., environmental exposures, diet), and evaluate potential joint effects of risk factors with genetic susceptibility.

Table 1. Characteristics of 17,471 NHL cases and 23,096 controls included in the InterLymph NHL Subtypes Project

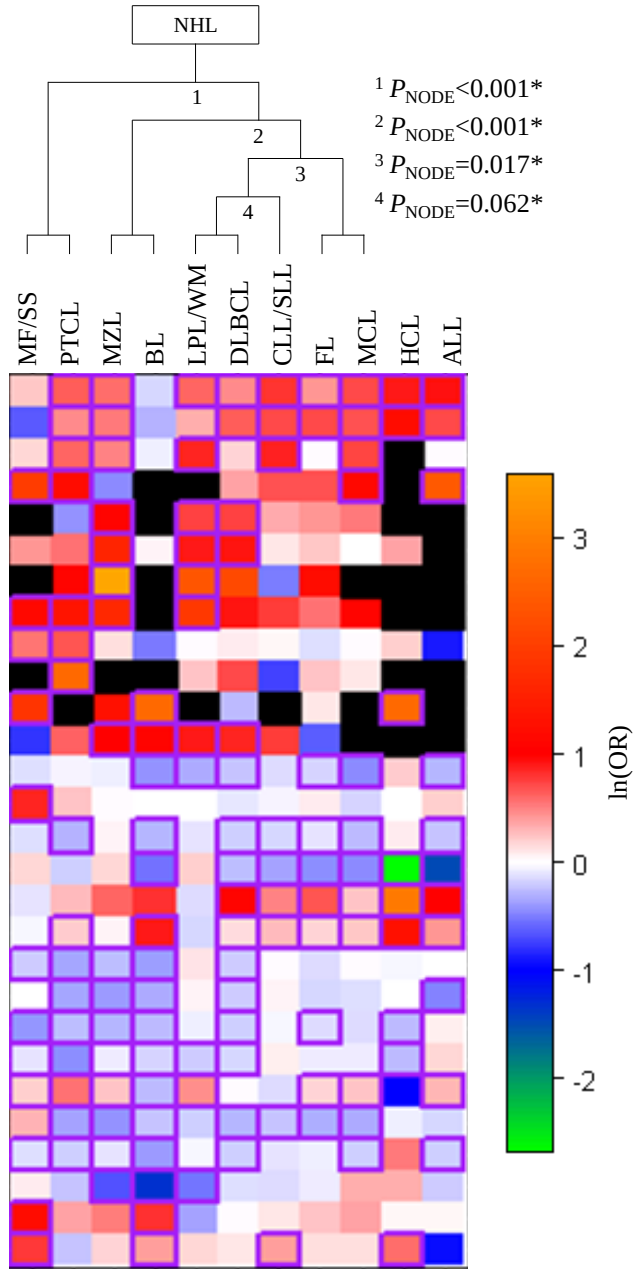
		Total NHL	Specified NHL subtypes*										
Characteristics	**Controls**	**cases**	**DLBCL**	**FL**	**CLL/SLL**	**MZL**	**PTCL**	**MCL**	**LPL/WM**	**MF/SS**	**BL**	**HCL**	**ALL**
Total No.	23,096	17,471	4667	3530	2440	1052	584	557	374	324	295	154	152
No. contributing studies	20	20	19	19	13	13	15	13	11	14	18	5	16
Population-based design (%)	77.3	80.2	81.4	82.4	67.9	80.5	80.8	78.1	77.8	86.4	83.7	70.8	68.4
By region (%)													
North America	49.6	45.9	44.1	52.5	36.1	53.3	40.6	45.4	41.7	61.4	62.4	0.0	36.8
Northern Europe	28.3	31.6	34.8	31.2	45.7	32.6	41.6	47.4	41.7	20.1	19.7	72.7	38.8
Southern Europe	19.0	18.4	16.2	9.2	17.0	8.3	15.1	3.2	9.4	17.3	16.6	20.8	21.1
Australia	3.0	4.0	4.9	7.1	1.2	5.8	2.7	3.9	7.2	1.2	1.4	6.5	3.3
Cases classified by WHO (%)	N/A	68.6	71.1	73.2	81.2	100	90.4	100	100	77.8	53.9	80.5	60.5
Male (%)	58.4	57.4	55.2	50.6	66.1	46.8	59.4	74.0	60.7	56.8	70.2	78.6	60.5
Non-Hispanic white (%)	93.4	91.5	90.4	91.5	95.7	87.2	87.7	93.9	94.1	83.6	84.4	96.1	86.8
Median age (years)[†]	59	60	59	58	64	61	56	62	64	56	53	55	41
(range)	(16-98)	(17-96)	(18-96)	(18-91)	(28-93)	(19-91)	(18-88)	(22-88)	(27-89)	(22-84)	(18-84)	(29-79)	(18-91)

Abbreviations: acute lymphoblastic leukemia/lymphoma (ALL); Burkitt/Burkitt-like lymphoma/leukemia (BL); chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL); diffuse large B-cell lymphoma (DLBCL); follicular lymphoma (FL); hairy cell leukemia (HCL); lymphoplasmacytic lymphoma/Waldenström macroglobulinemia (LPL/WM); mantle cell lymphoma (MCL); marginal zone lymphoma (MZL); mycosis fungoides/Sézary syndrome (MF/SS); non-Hodgkin lymphoma (NHL); peripheral T-cell lymphoma (PTCL); World Health Organization (WHO).

* We grouped cases into NHL subtypes according to the WHO classification [2, 39] using guidelines from the InterLymph Pathology Working Group [17, 40]. Total also includes rare subtypes with <100 cases and poorly specified subtypes (N=3342).

[†] Median age at diagnosis (cases) or interview (controls).

Figure 1. Family history, medical history, lifestyle, and occupational risk factors associated* with one or more NHL subtypes



Exposure category	Specific risk factor	P_{ASSET}†	$P_{homogeneity}$‡	Overall NHL OR (95% CI)§
Family history of hematologic malignancy	Any	1.62 10^{-22}	3.53 10^{-02}	1.72 (1.54 to 1.93)
	NHL	1.65 10^{-13}	0.5220	1.79 (1.51 to 2.13)
	Leukemia	1.34 10^{-11}	3.94 10^{-05}	1.51 (1.29 to 1.77)
	Multiple myeloma	7.52 10^{-04}	0.0223	1.77 (1.15 to 2.72)
	Hodgkin lymphoma	0.0020	0.4720	1.65 (1.18 to 2.29)
History of autoimmune disease	Any B-cell activating disease	3.80 10^{-22}	9.77 10^{-10}	1.96 (1.60 to 2.40)
	Sjögren's syndrome	6.30 10^{-18}	7.28 10^{-09}	7.52 (3.68 to 15.36)
	Systemic lupus erythematosus	1.92 10^{-08}	0.1840	2.83 (1.82 to 4.41)
	Any T-cell activating disease	0.0053	0.0117	1.07 (0.95 to 1.21)
	Celiac disease	5.18 10^{-11}	5.06 10^{-08}	1.77 (1.05 to 2.99)
	Systemic sclerosis/scleroderma	0.0051	0.0653	1.03 (0.41 to 2.58)
Hepatitis C virus seropositivity		2.32 10^{-08}	0.0021	1.81 (1.39 to 2.37)
History of atopic disease	Hayfever	9.06 10^{-09}	0.1180	0.82 (0.77 to 0.88)
	Eczema	5.01 10^{-05}	2.56 10^{-05}	1.01 (0.93 to 1.10)
	Allergy	5.87 10^{-05}	0.2390	0.86 (0.81 to 0.92)
History of blood transfusion	Transfusion occurring <1990	4.97 10^{-05}	0.0129	0.76 (0.67 to 0.87)
Anthropometric factors	Body-mass index as a young adult	4.20 10^{-09}	0.2810	1.95 (1.51 to 2.53)‖
	Height	0.0017	0.0241	1.20 (1.08 to 1.32)‖
History of alcohol consumption (≥1 drink per month)	Any alcohol	8.88 10^{-08}	0.0619	0.87 (0.81 to 0.93)
	Wine	4.87 10^{-09}	0.0139	0.85 (0.79 to 0.91)
	Liquor	4.08 10^{-06}	0.6570	0.84 (0.78 to 0.91)
	Beer	9.28 10^{-04}	0.1420	0.90 (0.84 to 0.97)
History of cigarette smoking	Duration of smoking	2.21 10^{-05}	3.19 10^{-09}	1.06 (0.99 to 1.14)‖
Recreational sun exposure		2.72 10^{-06}	0.7890	0.74 (0.66 to 0.83)‖
Socioeconomic status		3.36 10^{-05}	0.0610	0.88 (0.83 to 0.93)‖
Occupational history	Teacher	5.60 10^{-04}	0.0062	0.86 (0.77 to 0.95)
	Painter	0.0048	0.0846	1.22 (0.99 to 1.51)
	General farm worker	0.0082	0.3390	1.28 (1.10 to 1.50)

Abbreviations: acute lymphoblastic leukemia/lymphoma (ALL); Burkitt/Burkitt-like lymphoma/leukemia (BL); chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL); confidence interval (CI); diffuse large B-cell lymphoma (DLBCL); follicular lymphoma (FL); hairy cell leukemia (HCL); lymphoplasmacytic lymphoma/Waldenström macroglobulinemia (LPL/WM); mantle cell lymphoma (MCL); marginal zone lymphoma (MZL); mycosis fungoides/Sézary syndrome (MF/SS); non-Hodgkin lymphoma (NHL); odds ratio (OR); peripheral T-cell lymphoma (PTCL).

* Supplementary Table 2 provides the list of risk factors with P<0.05 in the logistic regression models separating each node.

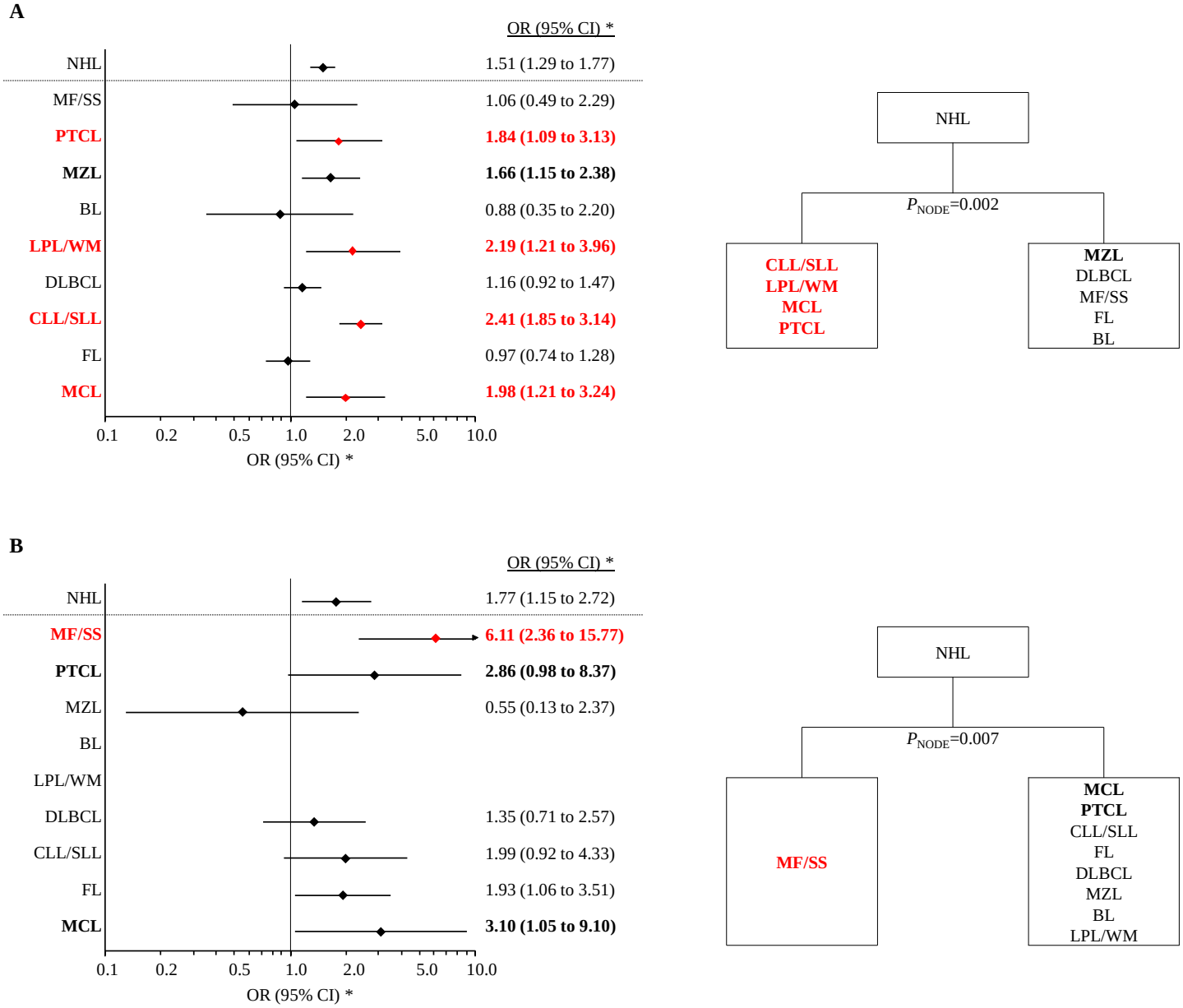
† Indicates P<0.01 from the subset-based statistical analysis (ASSET). For highly correlated variables (r>0.8; e.g., duration and pack-years of smoking), we selected the variable with the smaller PASSET. The only exception to this rule was history of beer consumption, for which the PASSET=2.77×10-04 for number of servings of beer per day, but "any history of beer consumption" (PASSET= 9.28 × 10-04) is presented to be consistent with other measures of alcohol consumption.

‡ Phomogeneity derived from the random effects meta-analysis Q statistic.

§ OR (95% CI) derived from fixed effects logistic regression, adjusting for age, race, sex, and study. For overall NHL, we used weighted logistic regression (using the function svyglm in R) where case status was defined by diagnosis of any NHL subtype, and sample weights were proportional to the prevalence of each subtype in the pooled study population divided by the prevalence in the United States (Supplementary Table 1).

‖ OR represents risk per increasing category of an ordinal variable for body-mass index as a young adult (<18.5, 18.5-22.4, 22.5-24.9, 25.0-29.9, ≥30 kg/m2), height (sex-specific quartiles, males: <172.0, 172.0-177.7, 177.8-181.9, ≥182.0 cm; females: <159.0, 159.0-162.9, 163.0-167.9, ≥168.0 cm), duration of cigarette smoking (0, 1-19, 20-29, 30-39, ≥40 years), recreational sun exposure (hours per week, study-specific quartiles available upon request), and socioeconomic status (low, medium, high).

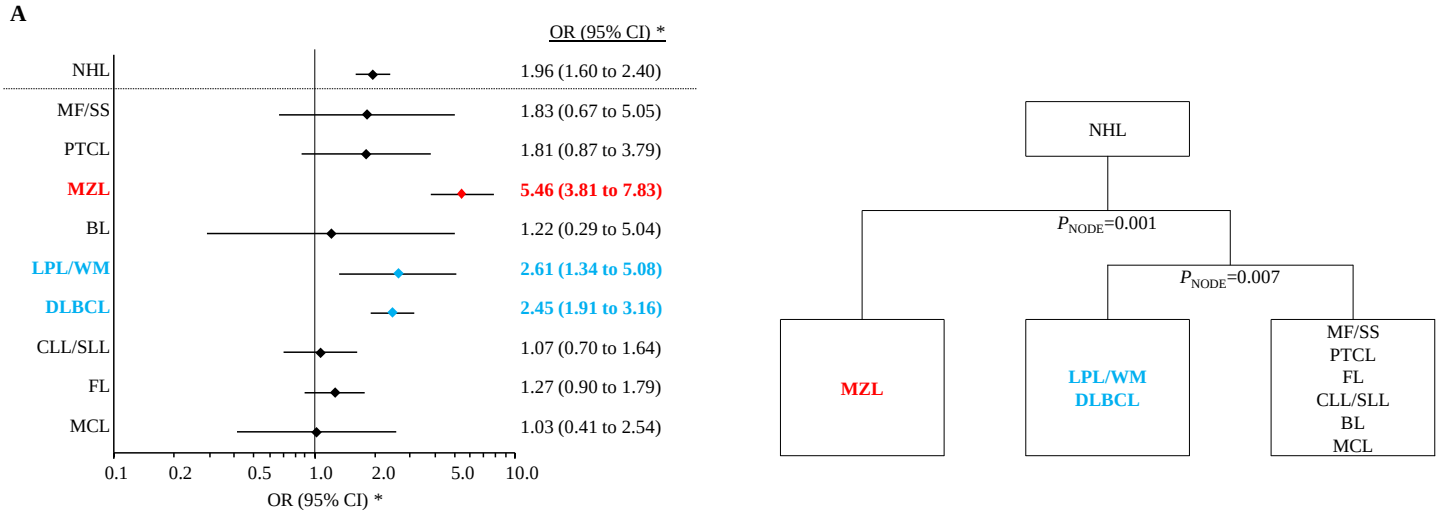
Figure 2. Risk of NHL overall and specific NHL subtypes associated with family history of leukemia (A) or multiple myeloma (B) in a first-degree relative.



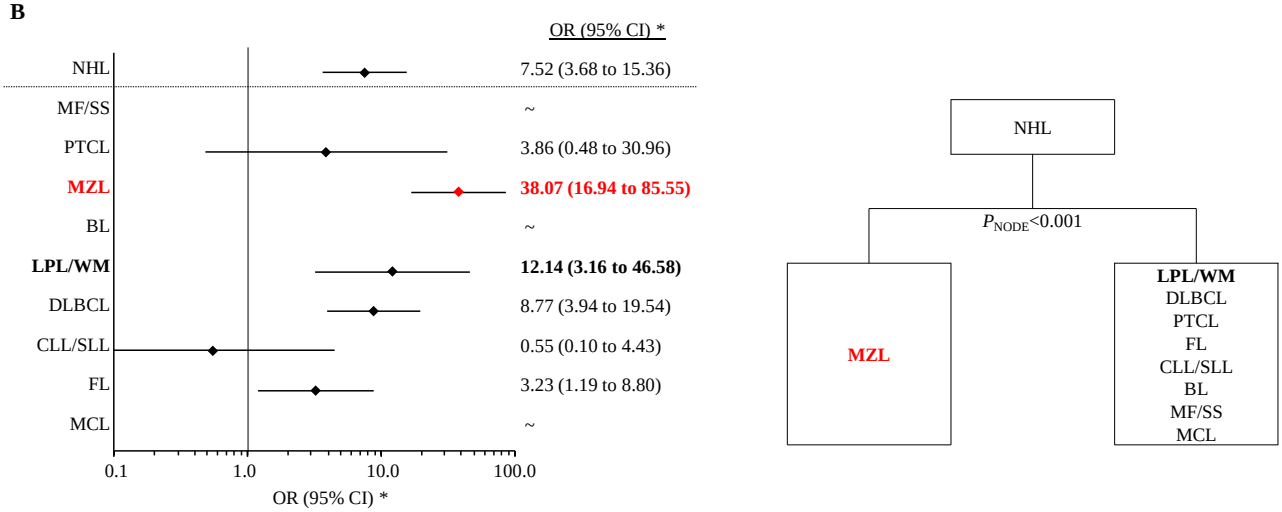
Abbreviations: Burkitt/Burkitt-like lymphoma/leukemia (BL); chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL); confidence interval (CI); diffuse large B-cell lymphoma (DLBCL); follicular lymphoma (FL); lymphoplasmacytic lymphoma/Waldenström macroglobulinemia (LPL/WM); mantle cell lymphoma (MCL); marginal zone lymphoma (MZL); mycosis fungoides/Sézary syndrome (MF/SS); non-Hodgkin lymphoma (NHL); odds ratio (OR); peripheral T-cell lymphoma (PTCL).
Bold font indicates the NHL subtypes identified as being associated in the subset-based statistical analysis (ASSET). Colors represent distinct nodes of the tree.
* OR (95% CI) derived from fixed effects logistic regression, adjusting for age, race, sex, and study. For overall NHL, we used weighted logistic regression (using the function svyglm in R) where case status was defined by diagnosis of any NHL subtype, and sample weights were proportional to the prevalence of each subtype in the pooled study population divided by the prevalence in the United States (Supplementary Table 1).

Figure 3. Risk of NHL overall and specific NHL subtypes associated with history of any B-cell activating autoimmune disease (A), Sjögren's syndrome (B), any T-cell activating autoimmune disease (C), and celiac disease (D).

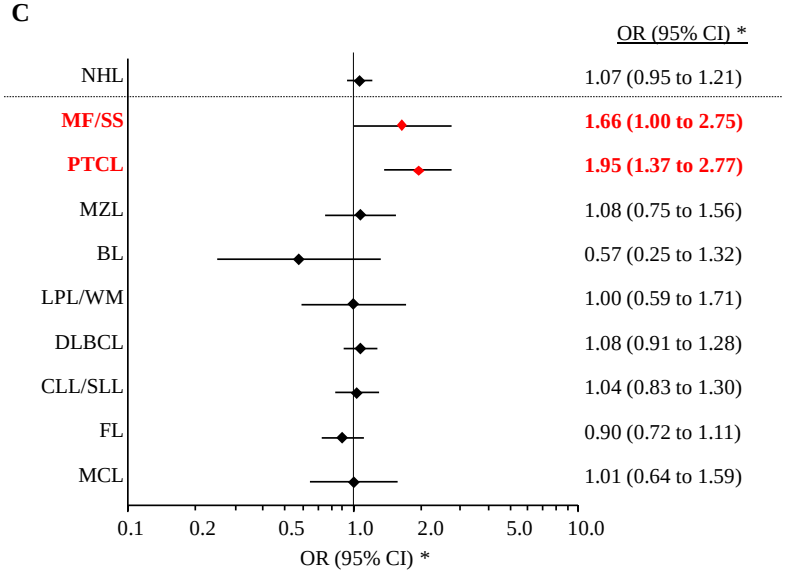
A

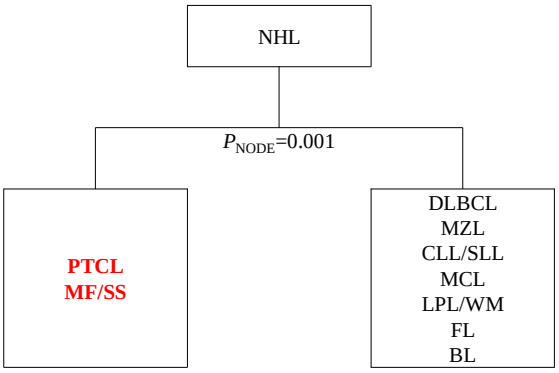


B

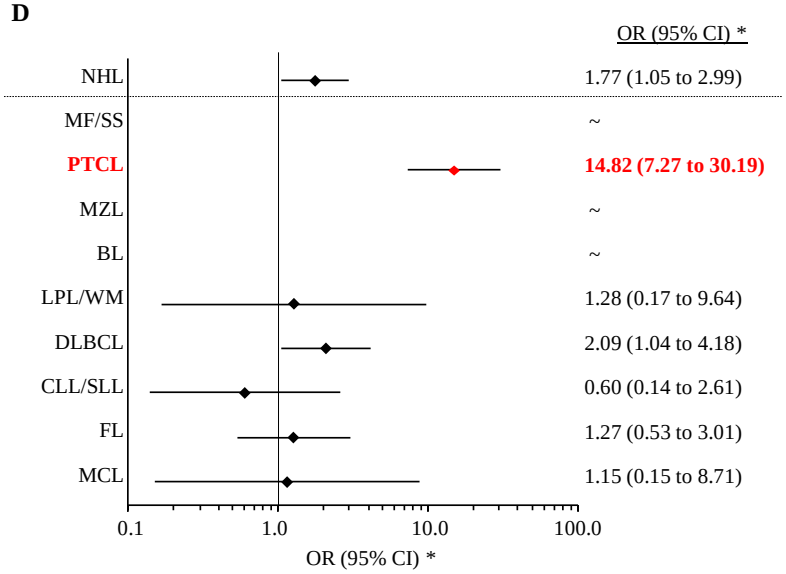


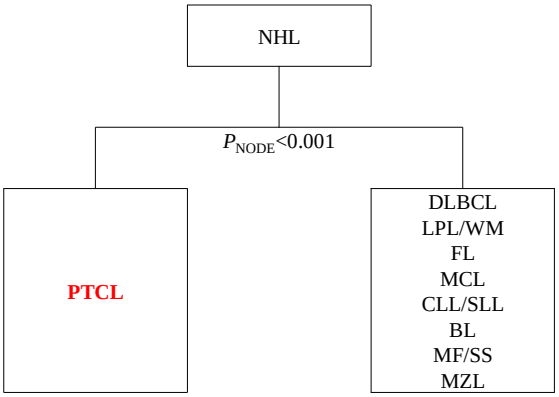
C





D

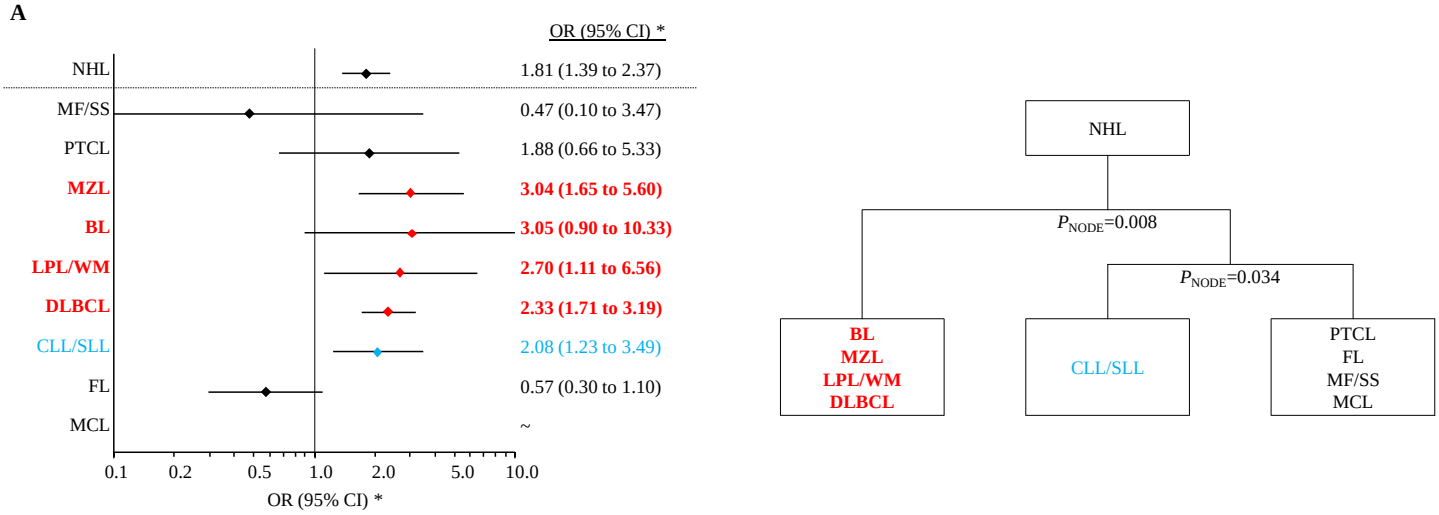




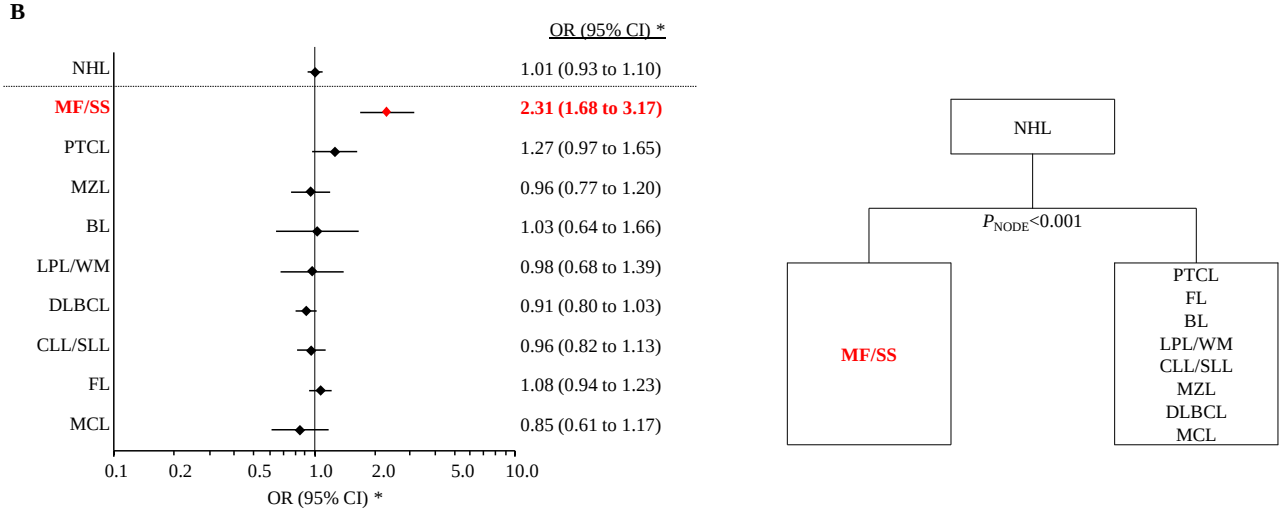
Abbreviations: Burkitt/Burkitt-like lymphoma/leukemia (BL); chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL); confidence interval (CI); diffuse large B-cell lymphoma (DLBCL); follicular lymphoma (FL); lymphoplasmacytic lymphoma/Waldenström macroglobulinemia (LPL/WM); mantle cell lymphoma (MCL); marginal zone lymphoma (MZL); mycosis fungoides/Sézary syndrome (MF/SS); non-Hodgkin lymphoma (NHL); odds ratio (OR); peripheral T-cell lymphoma (PTCL).
Bold font indicates the NHL subtypes identified as being associated in the subset-based statistical analysis (ASSET). Colors represent distinct nodes of the tree.
* OR (95% CI) derived from fixed effects logistic regression, adjusting for age, race, sex, and study. For overall NHL, we used weighted logistic regression (using the function svyglm in R) where case status was defined by diagnosis of any NHL subtype, and sample weights were proportional to the prevalence of each subtype in the pooled study population divided by the prevalence in the United States (Supplementary Table 1).

Figure 4. Risk of NHL overall and specific NHL subtypes associated with HCV seropositivity (A), eczema (B), and receipt of a blood transfusion prior to 1990 (C).

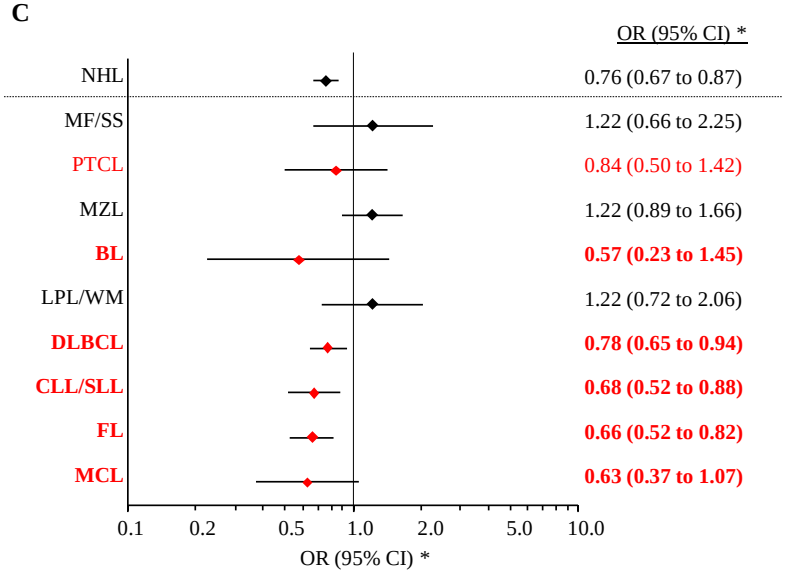
A

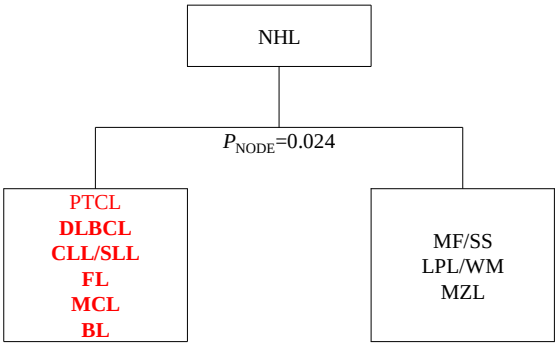


B



C

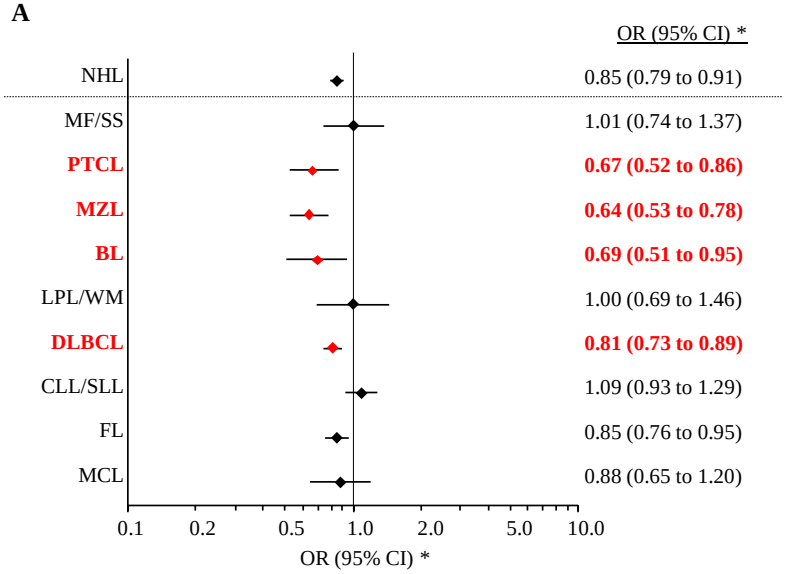


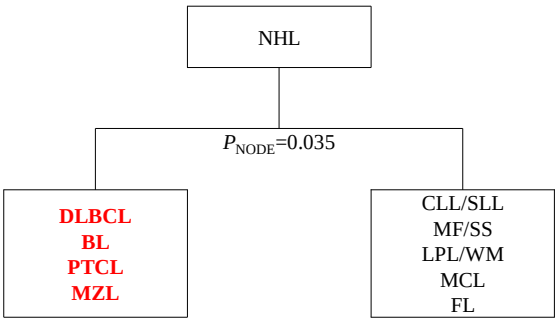


Abbreviations: Burkitt/Burkitt-like lymphoma/leukemia (BL); chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL); confidence interval (CI); diffuse large B-cell lymphoma (DLBCL); follicular lymphoma (FL); lymphoplasmacytic lymphoma/Waldenström macroglobulinemia (LPL/WM); mantle cell lymphoma (MCL); marginal zone lymphoma (MZL); mycosis fungoides/Sézary syndrome (MF/SS); non-Hodgkin lymphoma (NHL); odds ratio (OR); peripheral T-cell lymphoma (PTCL).
Bold font indicates the NHL subtypes identified as being associated in the subset-based statistical analysis (ASSET). Colors represent distinct nodes of the tree.
* OR (95% CI) derived from fixed effects logistic regression, adjusting for age, race, sex, and study. For overall NHL, we used weighted logistic regression (using the function svyglm in R) where case status was defined by diagnosis of any NHL subtype, and sample weights were proportional to the prevalence of each subtype in the pooled study population divided by the prevalence in the United States (Supplementary Table 1).

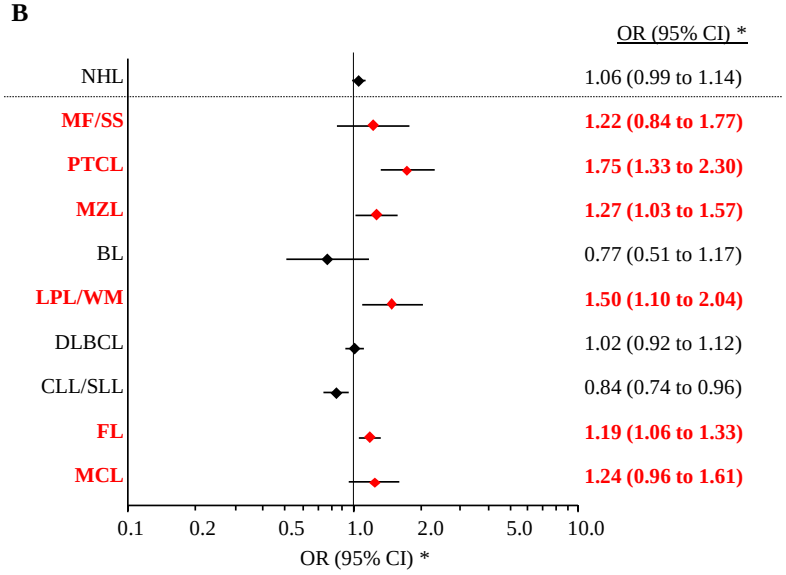
Figure 5. Risk of NHL overall and specific NHL subtypes associated with consumption of wine (A), duration of cigarette smoking* (B) and occupational history as a teacher (C) or painter (D).

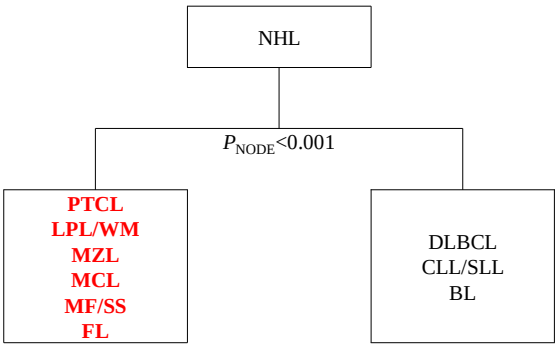
A



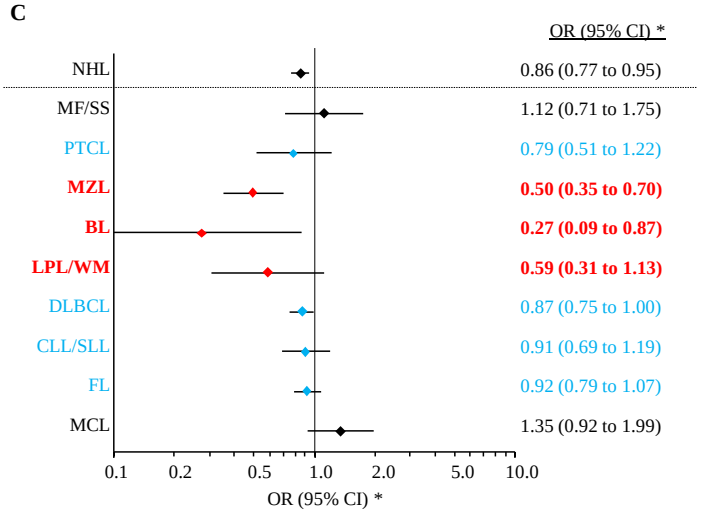


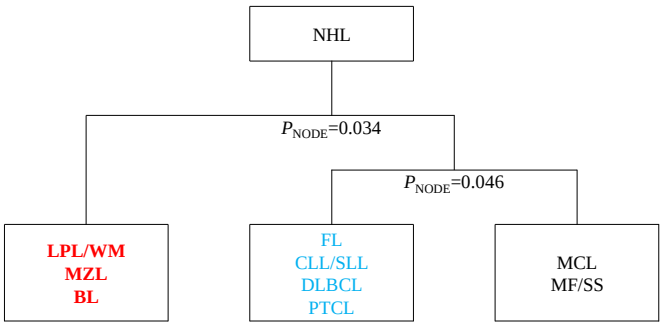
B



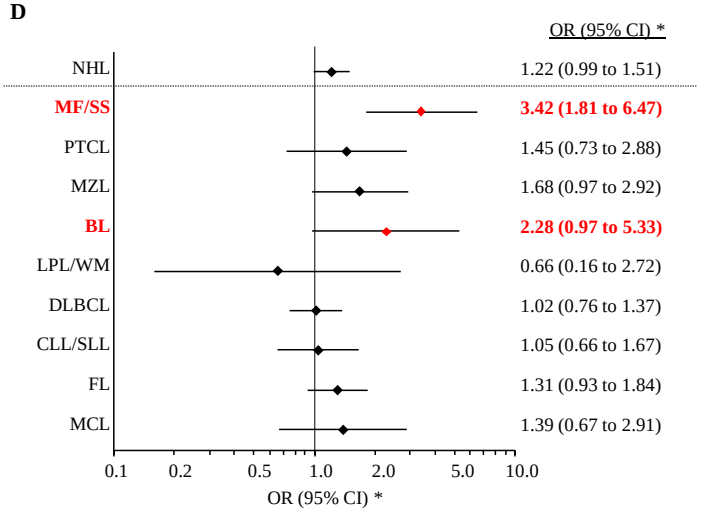


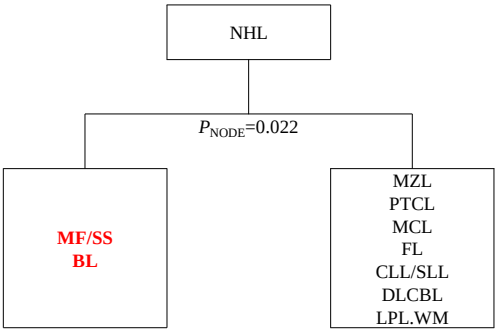
C





D





Abbreviations: Burkitt/Burkitt-like lymphoma/leukemia (BL); chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL); confidence interval (CI); diffuse large B-cell lymphoma (DLBCL); follicular lymphoma (FL); lymphoplasmacytic lymphoma/Waldenström macroglobulinemia (LPL/WM); mantle cell lymphoma (MCL); marginal zone lymphoma (MZL); mycosis fungoides/Sézary syndrome (MF/SS); non-Hodgkin lymphoma (NHL); odds ratio (OR); peripheral T-cell lymphoma (PTCL).
Bold font indicates the NHL subtypes identified as being associated in the subset-based statistical analysis (ASSET). Colors represent distinct nodes of the tree.
* OR (95% CI) derived from fixed effects logistic regression, adjusting for age, race, sex, and study. OR represents risk per increasing category of smoking duration (0, 1-19, 20-29, 30-39, ≥40 years).
For overall NHL, we used weighted logistic regression (using the function svyglm in R) where case status was defined by diagnosis of any NHL subtype, and sample weights were proportional to the prevalence of each subtype in the pooled study population divided by the prevalence in the United States (Supplementary Table 1).

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin 2012;62(1):10-29.
2. Jaffe ES, Harris NL, Stein H, *et al. World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon: IARC Press; 2001.
3. Groves FD, Linet MS, Travis LB, *et al.* Cancer surveillance series: non-Hodgkin's lymphoma incidence by histologic subtype in the United States from 1978 through 1995. J Natl Cancer Inst 2000;92(15):1240-1251.
4. Eltom MA, Jemal A, Mbulaiteye SM, *et al.* Trends in Kaposi's sarcoma and non-Hodgkin's lymphoma incidence in the United States from 1973 through 1998. J Natl Cancer Inst 2002;94(16):1204-1210.
5. Morton LM, Wang SS, Devesa SS, *et al.* Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. Blood 2006;107(1):265-276.
6. Manns A, Hisada M, Grenade LL. Human T-lymphotropic virus type I infection. Lancet 1999;353(9168):1951-1958.
7. Suarez F, Lortholary O, Hermine O, *et al.* Infection-associated lymphomas derived from marginal zone B cells: a model of antigen-driven lymphoproliferation. Blood 2006;107(8):3034-3044.
8. Cote TR, Biggar RJ, Rosenberg PS, *et al.* Non-Hodgkin's lymphoma among people with AIDS: incidence, presentation and public health burden. AIDS/Cancer Study Group. Int J Cancer 1997;73(5):645-650.
9. de Sanjose S, Benavente Y, Vajdic CM, *et al.* Hepatitis C and non-Hodgkin lymphoma among 4784 cases and 6269 controls from the International Lymphoma Epidemiology Consortium. Clin Gastroenterol Hepatol 2008;6(4):451-458.
10. Smedby KE, Vajdic CM, Falster M, *et al.* Autoimmune disorders and risk of non-Hodgkin lymphoma subtypes: a pooled analysis within the InterLymph Consortium. Blood 2008;111(8):4029-4038.
11. Wang SS, Slager SL, Brennan P, *et al.* Family history of hematopoietic malignancies and risk of non-Hodgkin lymphoma (NHL): a pooled analysis of 10 211 cases and 11 905 controls from

the International Lymphoma Epidemiology Consortium (InterLymph). Blood 2007;109(8):3479-3488.
12. Rothman N, Skibola CF, Wang SS, *et al.* Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium. Lancet Oncol 2006;7(1):27-38.
13. Skibola CF, Bracci PM, Nieters A, *et al.* Tumor necrosis factor (TNF) and lymphotoxin-{alpha} (LTA) polymorphisms and risk of non-Hodgkin lymphoma in the InterLymph Consortium. Am J Epidemiol 2010;171(3):267-276.
14. Morton LM, Zheng T, Holford TR, *et al.* Alcohol consumption and risk of non-Hodgkin lymphoma: a pooled analysis. Lancet Oncol 2005;6(7):469-476.
15. Kricker A, Armstrong BK, Hughes AM, *et al.* Personal sun exposure and risk of non-Hodgkin lymphoma: a pooled analysis from the InterLymph Consortium. Int J Cancer 2008;122(1):144-154.
16. Morton LM, Hartge P, Holford TR, *et al.* Cigarette smoking and risk of non-Hodgkin lymphoma: a pooled analysis from the InterLymph consortium Cancer Epidemiol Biomarkers Prev 2005;14(4):925-933.
17. Morton LM, Turner JJ, Cerhan JR, *et al.* Proposed classification of lymphoid neoplasms for epidemiologic research from the International Lymphoma Epidemiology Consortium (InterLymph). Blood 2007;110(2):695-708.
18. Willett EV, Morton LM, Hartge P, *et al.* Non-Hodgkin lymphoma and obesity: a pooled analysis from the InterLymph Consortium. Int J Cancer 2007;Epub ahead of print.
19. Spinelli JJ, Ng CH, Weber J-P, *et al.* Organochlorines and risk of non-Hodgkin lymphoma. Int J Cancer 2007;121(2767-2775).
20. Pottern LM, Linet M, Blair A, *et al.* Familial cancers associated with subtypes of leukemia and non-Hodgkin's lymphoma. Leuk Res 1991;15(5):305-14.
21. Hoar SK, Blair A, Holmes FF, *et al.* Agricultural herbicide use and risk of lymphoma and soft-tissue sarcoma. JAMA 1986;256:1141-1147.
22. Nelson RA, Levine AM, Marks G, *et al.* Alcohol, tobacco and recreational drug use and the risk of non-Hodgkin's lymphoma. Br J Cancer 1997;76(11):1532-1537.
23. Cerhan JR, Fredericksen ZS, Wang AH, *et al.* Design and validity of a clinic-based case-control study on the molecular epidemiology of lymphoma. Int J Mol Epidemiol Genet 2011;2(2):95-113.

24. Cerhan JR, Fredericksen ZS, Novak AJ, *et al.* A two-stage evaluation of genetic variation in immune and inflammation genes with risk of non-Hodgkin lymphoma identifies new susceptibility locus in 6p21.3 region. Cancer Epidemiol Biomarkers Prev 2012;21(10):1799-806.
25. Chatterjee N, Hartge P, Cerhan JR, *et al.* Risk of non-Hodgkin's lymphoma and family history of lymphatic, hematologic, and other cancers. Cancer Epidemiol Biomarkers Prev 2004;13(9):1415-1421.
26. Zahm SH, Weisenburger DD, Babbitt PA, *et al.* A case control study of non-Hodgkin's lymphoma and the herbicide 2,4-dichlorophenoxyacetic acid (2,4-D) in eastern Nebraska. Epidemiology 1990;1:349-356.
27. Chiu BCH, Kolar C, Gapstur SM, *et al.* Association of NAT and GST polymorphisms with non-Hodgkin's lymphoma: a population-based case-control study. Br J Haematol 2005;128(5):610-615.
28. Holly EA, Lele C, Bracci PM, *et al.* Case-control study of non-Hodgkin's lymphoma among women and heterosexual men in the San Francisco Bay Area, California. Am J Epidemiol 1999;150(4):375-389.
29. Kelly JL, Friedberg JW, Calvi LM, *et al.* A case-control study of ultraviolet radiation exposure, vitamin D, and lymphoma risk in adults. Cancer Causes Control 2010;21(8):1265-1275.
30. Morton LM, Holford TR, Leaderer B, *et al.* Alcohol use and risk of non-Hodgkin's lymphoma among Connecticut women (United States). Cancer Causes Control 2003;14(7):687-94.
31. Monnereau A, Orsi L, Troussard X, *et al.* History of infections and vaccinations and risk of lymphoid neoplasms: Does influenza immunization reduce the risk? Leukemia 2007;21(9):2075-2079.
32. Besson H, Brennan P, Becker N, *et al.* Tobacco smoking, alcohol drinking and non-Hodgkin's lymphoma: A European multicenter case-control study (Epilymph). Int J Cancer 2006;119(4):901-8.
33. Vineis P, Crosignani P, Sacerdote C, *et al.* Haematopoietic cancer and medical history: A multicentre case control study. J Epidemiol Community Health 2000;54(6):431-436.
34. Franceschi S, Serraino D, Bidoli E, *et al.* The epidemiology of non-Hodgkin's lymphoma in the north-east of Italy: A hospital-based case-control study. Leuk Res 1989;13(6):465-472.

35. Talamini R, Montella M, Crovatto M, *et al.* Non-Hodgkin's lymphoma and hepatitis C virus: a case-control study from northern and southern Italy. Int J Cancer 2004;110(3):380-385.
36. Smedby KE, Hjalgrim H, Melbye M, *et al.* Ultraviolet radiation exposure and risk of malignant lymphomas. J Natl Cancer Inst 2005;97(3):199-209.
37. Willett EV, Smith AG, Dovey GJ, *et al.* Tobacco and alcohol consumption and the risk of non-Hodgkin lymphoma. Cancer Causes Control 2004;15(8):771-780.
38. Hughes AM, Armstrong BK, Vajdic CM, *et al.* Sun exposure may protect against non-Hodgkin lymphoma: A case-control study. Int J Cancer 2004;112(865-871).
39. Swerdlow SH, Campo E, Harris NL, *et al. World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon: IARC Press; 2008.
40. Turner JJ, Morton LM, Linet MS, *et al.* InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. Blood 2010;116(20):e90-8.

2012 年 10 月

中国儿童与家庭队列研究(CFCS)

数据管理和质控研讨会
参考指南

接收方

美国国家癌症研究所
中国疾病预防控制中心
美国疾病预防控制中心

提交方

RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709

RTI 项目编号 0212208.001.220.001

RTI 项目编号
0212208.001.220.001

中国儿童与家庭队列研究(CFCS)

数据管理和质控研讨会
参考指南

2012 年 10 月

接收方

美国国家癌症研究所
中国疾病预防控制中心
美国疾病预防控制中心

提交方

RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709

版本控制表

版号	日期	负责人	记录
0.1	09/24/2012	Mandy Sha	提交供内部审阅的英文版草案
0.2	09/28/2012	Kristen Stolka	向 NCI 提交英文版草案
0.3	10/5/2012	Kristen Stolka	向 NCI 提交经过修订的英文版草案
1.0	10/10/2012	Mandy Sha	向 China CDC 提交中文版
1.1	10/20/2012	Mandy Sha	向 China CDC 提交修订的中文版及英文版

术语表

调研工具 （Instrument）：用于收集数据的工具，可以是调查问卷，记录，或表格的形式。

数据追踪收集：是在现场调研数据收集工作完成之后，再试着收集在调研现场没收集到的、或收集不完整的数据信息。

个案：个案，是数据收集工作中，用作分析的单元。一个个案可以与多种表格和调查问卷都有关。在 CFCS 任务 1 预试验中，一个“家庭组”就是一个个案。在任务 2 预试验中，“母亲和她的孩子组成的母子对”则被认为是一个个案。

变量：是调研工具中，含有某种资料信息的单元。一个个案，可能有许多变量；调查问卷中一个问题，可以被转换成一个或多个变量。

变量名称：变量的简短名称。

变量类型：变量的属性，如字符型（文字）或数字型，取决于该变量的赋值。

变量赋值：为数字（如 1,2,3，...）或字母序列（即，A，B，C ...）。在变量中，表示某种资料信息。

变量赋值标签：标签是以人性化的方式，描述和记录某个变量的内容及变量的数据值。例如，变量赋值 1 可能会得到一个“有”的标签。

变量宽度：变量赋值的显示所需要的宽度。譬如，变量赋值 1 的宽度是 1 个按键宽度，变量赋值 11 则是 2 个按键宽度。

编码本（数据字典）：总汇所有与某数据库相关的信息，包括变量名称，变量赋值和它们的说明。

Contents

Section **Page**

1. 简介 **1**

1.1 如何使用本指南 1

2. 接收记录和分组流程 **3**

2.1 接收记录 3

2.2 分组流程 5

3. 人工数据审阅和编辑 **9**

3.1 集中审阅问卷及表格 9

3.2 人工数据编辑 12

3.3 创建编码本或数据字典 20

4. 数据输入和验证 **26**

4.1 数据输入 27

4.2 数据验证 29

5. 数据清理和数据库的递交 **29**

5.1 数据清理 30

5.1.1 制作所有数据库内容的频率分布 31

5.1.2 使用散点图和流程图 35

5.1.3 交叉制表 (Cross-tabulations) 38

5.1.4 重要提示 40

6. 数据质量和管理的其他问题 **45**

Figures

Number **Page**

图 1-1 (Figure 1-1). 数据质控和管理周期 2
图 2-1 (Figure 2-1). 数据管理过程中的接收记录和分组流程 3
图 2-2 (Figure 2-2). 电子收集记录系统 4
图 2-3.(Figure 2-3). 通用的小组封面 6
图 2-4.(Figure 2-4). 为具体项目量身定制的小组封面样品 8
图 3-1 (Figure 3-1). 数据管理周期中的人工数据审阅和编辑 9
图 3-2 (Figure 3-2). 关于教育程度的矛盾答案 13
图 3-3 (Figure 3-3). 关于经常性的矛盾答案 13
图 3-4 (Figure 3-4). 关于居住问题的矛盾回答 14
图 3-5 (Figure 3-5). 人工数据审阅和编辑的调查问卷 17
图 4-1 (Figure 4-1). 在数据管理过程中的数据输入和验证的流程 26
图 4-2 (Figure 4-2). 用选择数字来回答问题 27
图 4-3 (Figure 4-3). 通过填写数字来回答问题 27
图 4-4 (Figure 4-4). 通过书面文字回答问题。 28
图 4-5 (Figure 4-5). EpiInfo 的范围检查示例 28
图 5.1 (Figure 5-1). 数据管理过程中的数据清理和数据库的递交 30
图 5-2 (Figure 5-2). ID 数字样本的打印件,显示一个重复和错误的数据输入 32
图 5-3 (Figure 5-3). 用来评估跳转问题回答方式的调查分布频率 33
图 5-4 (Figure 5-4). 数据以竖列形式打印出来的样本示例 34
图 5-5 (Figure 5-5). 数据编码本与数据之间对于变量和变量标签的偏差 34
图 5-6 (Figure 5-6). 男女身高和日进卡路里的散点图 36
图 5-7 (Figure 5-7). CFCS 任务 1 预试验孩子问卷调查的流程图 (C007 – C012) 37
图 5-8 (Figure 5-8). C007 和 C008A 的交叉列表 38
图 5-9 (Figure 5-9). C008A 和 C008B 的交叉列表 39
图 5-10 (Figure 5-10). C012 和 C007 的交叉列表 39
图 5-11 (Figure 5-11). 数个变量的交叉列表 （C007, C008A, COO8B,和 C012） 40
图 5-12 (Figure 5-12). 用来评估缺省数据处理方法的调查问卷频率分布 42

Tables

Number **Page**

表 2-1 (Table 2-1). 美国儿童教育研究项目所用的接收编码..... 5
表 3-1 (Table 3-1). CFCS 任务 1 预试验问题（范例）.....11
表 3-2 (Table 3-2). 膳食日记和食物频率问卷的人工审阅核对表用途.....12
表 3-3 (Table 3-3). RTI 常用的一致性代码.....15
表 3-4 (Table 3-4). 编辑问题记录表.....16
表 3-5 (Table 3-5). 用于图 3-5 的问卷编码本.....17
图 3-6 (Table 3-6). 三天紫外线日照记录（范例）.....23
图 3-7 (Table 3-7). 三天紫外线日照记录编码本.....25

1. 简介

本指南具体列出将在 2012 年十月北京举行的数据管理和质控研讨会的内容。这个研讨会涵盖手写问卷数据质量控制和管理的整个过程，也为中国儿童与家庭队列研究(Children and Families Cohort Study,简称 CFCS)任务 1 及任务 2 预试验量身定做。图 1-1(Figure 1-1)展示了整个过程,包括:

- 数据接收和分组流程（第 2 章）
- 指南数据查阅和整理（第 3 章）
- 数据输入和验证（第 4 章）
- 数据清理和数据库递交（第 5 章）
- 文件整理和报告(整个过程中)
- 对实体文档资料(如纸质文件)的安全存放,检索,和处理（整个过程中）

每个章节都列出了在质量控制和管理过程中为提高数据质量而应采取的一系列步骤。对于如何完成这些相干任务的详细指示都通过实例来展示,这些例子来自于美国以往的研究项目和 CFCS 任务 1 及任务 2 预试验。所有展示的数据都不包含识别个人的信息。第 6 章涵盖其他数据质量方面的问题。

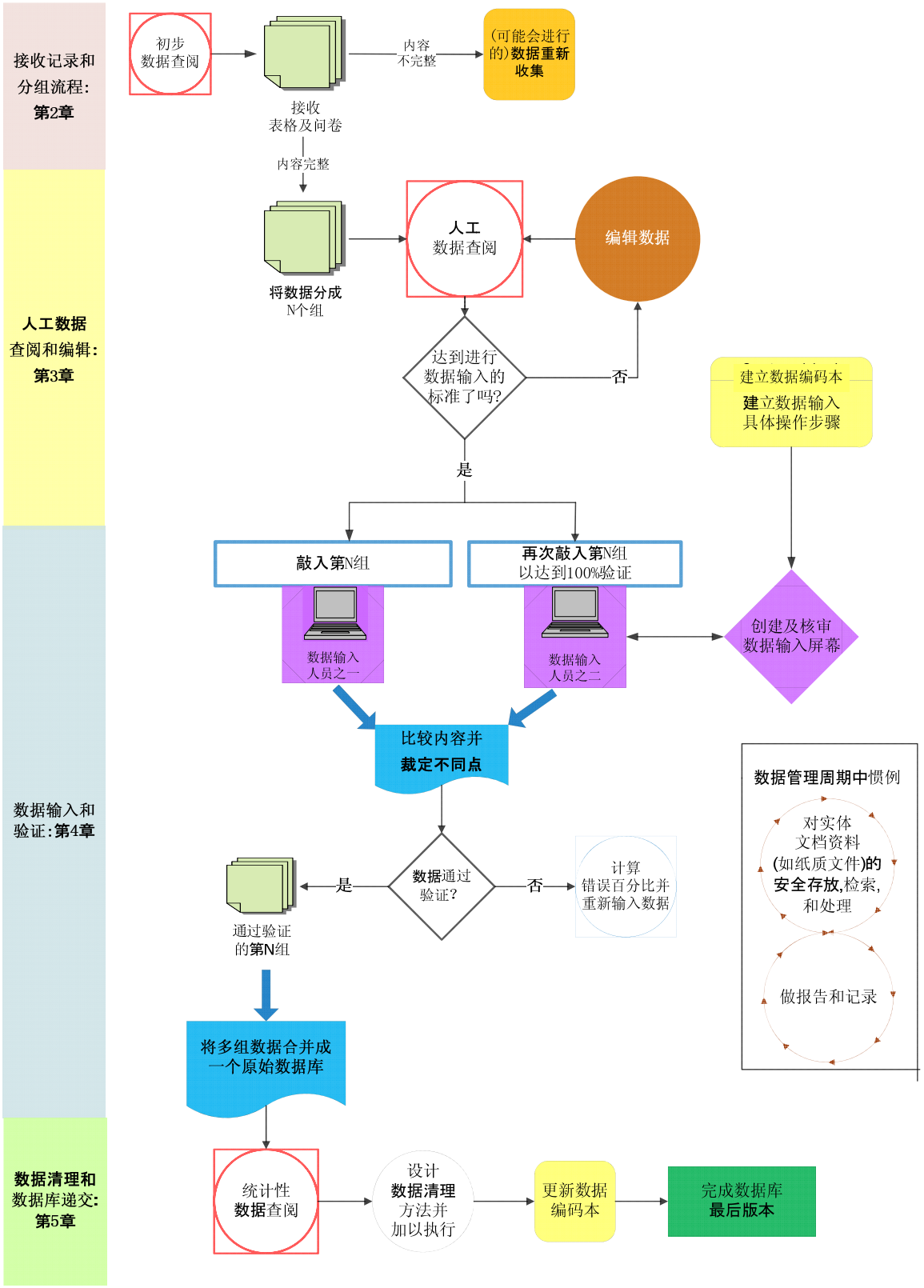
本指南中的方法都经美国的 RTI International (RTI) 实践证明，即能保证效率，又能保证数据质量。

1.1 如何使用本指南

本指南是为用来辅助参加研讨会的人员。当用本指南为长期参照的材料,以下列出使用建议。

1. 请熟悉本指南前面词汇表中的名词。
2. （首次阅读本指南的读者）请将本指南从前至后一次读完,了解前一章所描述内容,以更好理解下面一章中的内容。
3. （再次阅读本指南的读者）用目录来找出与你工作内容相关的部分以便仔细阅读。
4. 完成章节后面提供的关于 CFCS 项目的练习(“练习”)
5. 有些部分可能因为 CFCS 项目以后的需要而有所变动及更新。

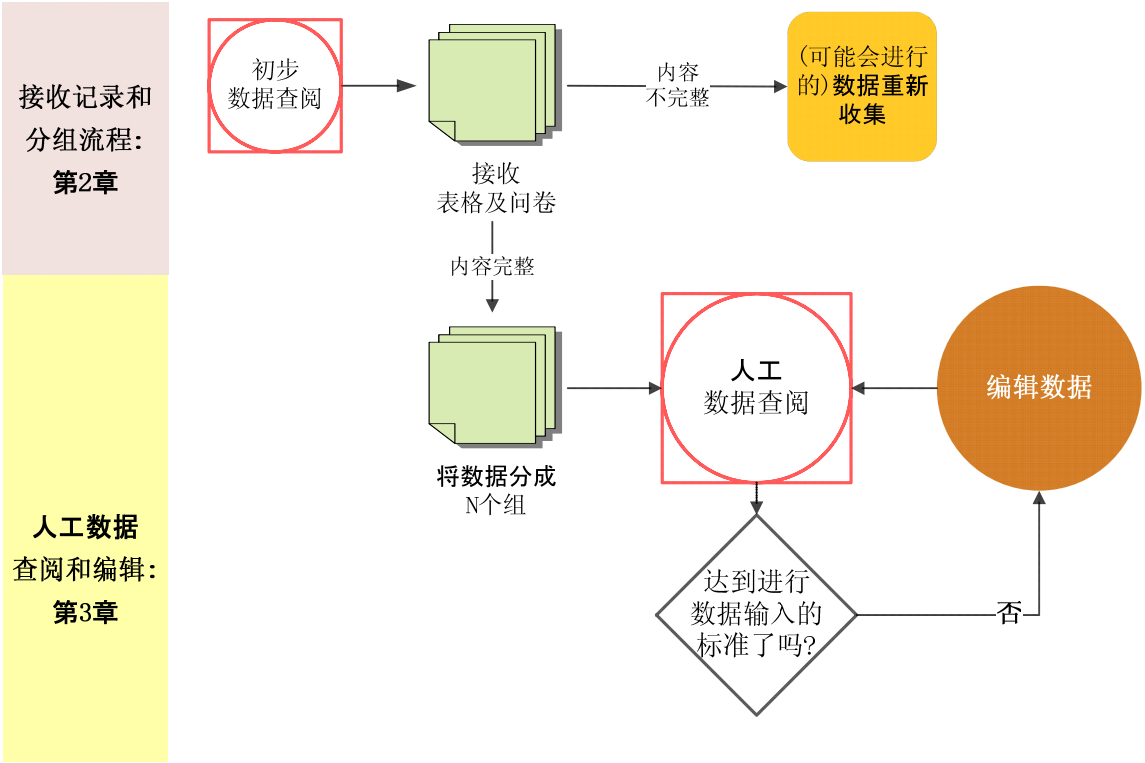
图 1-1 (Figure 1-1). 数据质控和管理周期



2. 接收记录和分组流程

接收记录和分组流程这两个环节的目的是由数据管理人员集中记录和整理调查问卷和表格。图 2-1 显示收集记录和分组流程是所有数据管理整个过程的开始。当调查问卷和表格从调研地收集上来时,数据管理人员要先对这些材料进行初步审阅,将完整的问卷和表格进行分组,以便人工查阅和编辑。若是不完整，可加以追踪收集遗漏或没完成的数据。

图 2-1 (Figure 2-1). 数据管理过程中的接收记录和分组流程



2.1 接收记录

接收记录是指数据管理工作人员对在调研地收集上来的调查问卷和表格进行记录的过程。这些调查问卷和表格可以是直接从调查对象而来或由调查员收集来的。

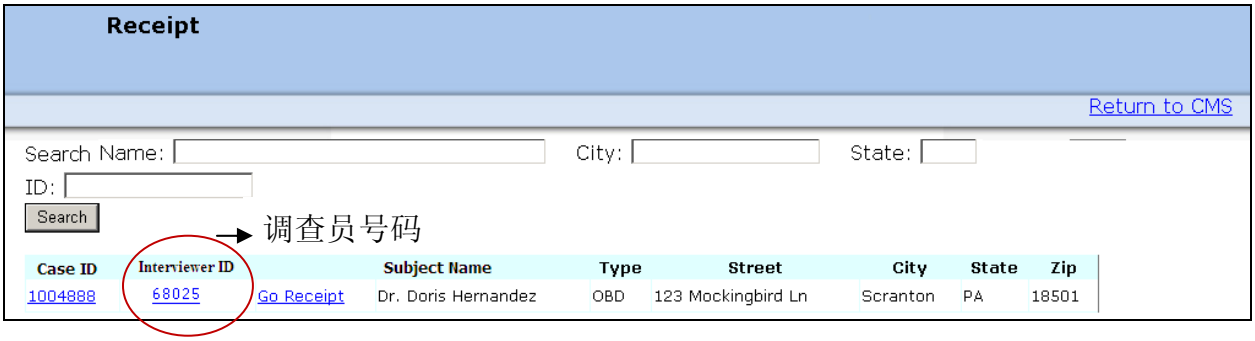
接收记录之前,应该对 CFCS 任务 1 预试验的“家庭组”和任务 2 预试验“母子对”中的所有问卷,表格,和日记进行初步的查阅。这个步骤是用来决定是否所有跟受试对象有关的问卷和表格是完整的，以及所有重要的变量数据都已收集到。初步的查阅只集中在重要的变量。例如,某些问题绝不可缺漏回答，或者得记录成特定的方式,才能让这个调查问卷达到操作意义上的数据收集“完成”。有的研究项目要求所有问卷中的问题都回答才能达到操作意义上的“完成”状态。而有些项目,即使某些问题或者表格本身有所遗漏,仍可以被看做是操作意义上的数据收集“完成”。 “完成”的定义应由研究项目的领导层来决定。

当一个调查问卷或表格被定义为“未完成”状态，数据管理员应当和项目领导来讨论决定是否需要做数据追踪收集。数据追踪收集是指对遗漏或不完整的数据再次或重新进行采集。如果数据追踪收集不成功或者决定不进行这一步骤，这个表格的收集记录应当用收据编码来注明。

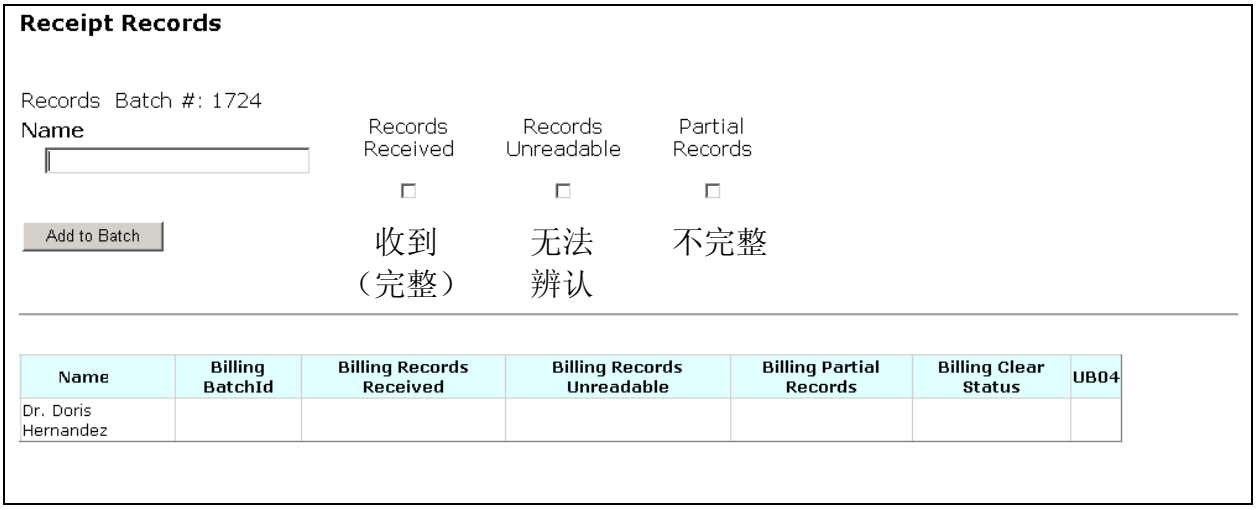
收据编码是为项目而专门制定的。图 2-2(Figure 2-2)展示了在美国一个关于病人帐单研究的电子收集记录系统。为更好跟踪接收状态，使用这类电子接受系统会很有帮助。图 2-2 显示，当纸质的病人帐单从调研地采集上来后，数据管理员在电子收集记录系统中找到相关的个案编号，点选，然后在相应的方框里勾选此个案的收集状态。这个项目有 3 个收据编码：收到（完整），无法辨认，或者不完整。 注意调查员号码(数据采集员的编号)是记录中的一部分，这样是为了能方便找出发送人是谁。

图 2-2 (Figure 2-2). 电子收集记录系统

查找和点选个案:



记录接收编码:



记录接收编码有利于定期报告,因为可以量化编码来告知个调研工具接收状态的细节情况。表 2-1(Table 2-1)显示了另一个美国的项目。这是一个关于儿童教育的研究项目。因为项目需要，所用的收集编码也不同。

表 2-1 (Table 2-1). 美国儿童教育研究项目所用的接收编码

接收编码	
收到—完成	未收到—父母拒绝
收到—空白	未收到—儿童拒绝
收到—有问题	未收到—丢失

总之,数据管理人员要首先进行初步审阅,来决定以下几个问题(以任务 1 预试验“家庭组”的问卷交到数据管理管理中心为例):

1. 所有关于这个“家庭组”的调研工具都收上来了吗?
2. 所有(由项目领导者来决定)重要的变量都收集到了吗?
3. 这个“家庭组”的数据都属于操作意义上的“完成”状态吗?

2.2 分组流程

分组流程是指数据管理人员将一定数量的签收到的调查问卷和表格分在同一组的操作流程。分组的目的，是为了在数据管理过程中，更好地组织和跟踪的实体文档资料(如纸质问卷，录音带等)--尤其是当由于需要进行审阅，或由于其他的原因，需将某一表格单独提取出来时。更重要的是，分组提高了效率，因为同一组里的问卷可在同一时间经历其数据管理周期中同一步骤。同时，分组使得处理这些表格变得更容易一些。数据管理人员和项目负责人确定每个组的结构组成。 例如，应该把哪些个案和调查工具分在一组。小组的大小，往往取决于什么可以一次被检阅或处理，如多少的问卷可以扎成一捆。每种问卷最好照调查工具种类以及个案编码顺序排列，以便找寻。此外，分成多个小组，可以让写工作进展报告或计划时更容易一些，因为每个组只有较少的表格，处理类似的表格可以更快。若是预试验或小型项目，数据管理人员可能决定不另外分组，一个小组即可。

为了记录在数据管理过程中各个小组都在哪里，每个小组都配给了一个组号，并有了一个特定的封面，称为小组封面。

图 2-3(Figure 2-3)是一个通用的小组封面，用于在分组后的数据管理流程中（如人工数据审阅，编辑，和数据输入流程中）跟踪该组的各个文件，直到该组所有的文件资料都被归档。这将使处理该组文件的任何人，都可以清楚地了解这些文件应该通过的，和已经通过的任何一道程序。小组封面也可以作为一个报告工具。即使在文件资料归档/备案以后，如因任何原因要提取文件，可在此小组封面上，注明文件的提取人和接收人。

正如在图 2-3 中示出的，小组封面上有一个部分，叫小组发送纪录，包括以下的内容：

- (A) 任务/步骤说明：这部分列出该组文件应经过的所有研究步骤
- (B) 有个标明每一个的任务/步骤是否完成的地方
- (C) 有个标明每一个的任务/步骤完成的日期的地方
- (D) 有个标明完成每一个的任务/步骤的人的名字的缩写的地方
- (E) 有个标明该组文件的负责人的姓名（此人在分组后负责此组）的地方

图 2-3.(Figure 2-3). 通用的小组封面

BATCH COVER SHEET

Project Name
Form Name
RTI Project #

BATCH NUMBER: 组号

NUMBER OF DOCUMENTS: _____

CASE ID#

BATCH ROUTING 小组发送纪录

A DESCRIPTION OF TASK	(√)	DATE	INITIAL	CUSTODY RECEIPT
A. BATCH EVENT	____	__/__/__	_____	_____
B. TRANSFER TO EDIT / CODING	B	C	D	E
C. EDIT / CODE	____	__/__/__	_____	_____
D. EDIT CODE QC	____	__/__/__	_____	_____
E. PASS EDIT EVENT	____	__/__/__	_____	_____
F. TRANSFER TO KEYING	____	__/__/__	_____	_____
G. KEYING	____	__/__/__	_____	_____
H. KEYING QC	____	__/__/__	_____	_____
I. TRANSFER TO DOCUMENT CONTROL	____	__/__/__	_____	_____
J. STORAGE	____	__/__/__	_____	_____

DOCUMENT CONTROL LOG 控制纪录

DATE:	__/__/__	TO:	_____	BY:	_____	PURPOSE:	_____	DATE: __/__/__	REC'D BY: _____
DATE:	F	TO:	G	BY:	H	PURPOSE:	I	DATE: J	REC'D BY: K
DATE:	__/__/__	TO:	_____	BY:	_____	PURPOSE:	_____	DATE: __/__/__	REC'D BY: _____
DATE:	__/__/__	TO:	_____	BY:	_____	PURPOSE:	_____	DATE: __/__/__	REC'D BY: _____
DATE:	__/__/__	TO:	_____	BY:	_____	PURPOSE:	_____	DATE: __/__/__	REC'D BY: _____
DATE:	__/__/__	TO:	_____	BY:	_____	PURPOSE:	_____	DATE: __/__/__	REC'D BY: _____

另一个部分是该组文件的控制记录。虽然这部分可能不经常用，但也非常重要，尤其是当此组文件在归档后，又被调出。控制记录部分可包括以下内容：

- (F) 此组文件被调出的日期
- (G) 有个标明此组文件的提取人的地方
- (H) 有个标明此组文件的接收人（此时的文件责任人）的地方
- (I) 有个标明此组文件组调出的原因或目的的地方
- (J) 标明此组文件归档的日期的地方
- (K) 标明此组文件的接收人（此时的文件责任人）的地方

必须注意的是：小组封面可以、并且应该为具体项目量身定制。图 2-4(Figure 2-4) 显示了 RTI 的一研究项目所用的小组封面表。它列出下列内容：组号（114），分组人（CW），它被分组的日期（2012 年 9 月 18 日），在此组中所有的个案编号（CASE ID）且顺序排列，文件类型（问卷），特定的任务(，任务完成日期，和完成任务的人的名字的缩写。它还有一个控制记录，用于管理此组文件及其个案文件。

图 2-4.(Figure 2-4). 为具体项目量身定制的小组封面样品

组号 114 **小组发送记录表**

任务	完成	日期	签名
a. 分组	☑	9-18-12	CW
b. 移至到扫描	☑	9-18-12	CW
c. 扫描	☑	9-12-12	CW
d. 移至到邮寄处	☐		
e. 邮寄给客户	☐		

文件类型: 问卷

个案号	发出记录			归还记录		
	接收人	提取人	日期	接收人	提取人	日期
600706						
601721						
602237						
602352						
602363						
602404						
604352						
605135						
605848						
602413						
602814						
602816						
602821						
603358						
603552						

小组文件数目: 15

小组文件控制表

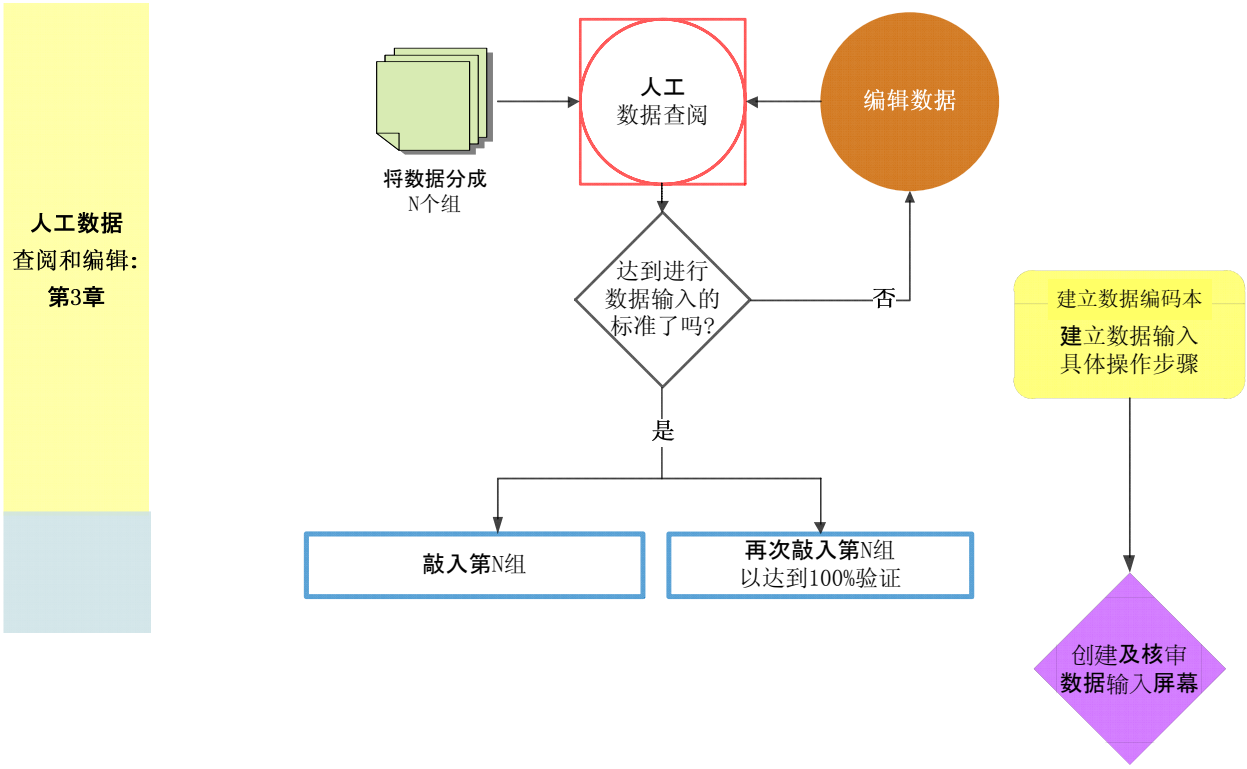
小组文件发出记录			目的	小组文件归还记录		
接收人	提取人	日期		接收人	提取人	日期

3. 人工数据审阅和编辑

人工数据审阅和编辑的目的，是让数据的管理人员、研究专题专家和项目负责人，对收集的数据集中彻底地质量检查。是在数据输入建立电子资料库前所做的步骤。并将所发现的错误都改掉。

图 3-1(Figure 3-1)是每组调查问卷和表格通过人工数据审阅的流程。如果审阅确定该组调查问卷和表格是完整的，正確的，可以被键入了，即可将实体文档资料(如纸质问卷)送去进行数据输入。如果该组调查问卷和表格需要做数据追踪收集（譬如从调查员那里取得遗漏的资料）或者需要校正（此步骤叫数据编辑）,则不可进行数据输入。先要根据编辑规则，对问卷答案进行编辑，然后再被送去进行数据输入。数据输入开始之前，先要建立编码本，然后根据此编码本，制定出数据输入规则（给数据输入人员的工作指示）。

图 3-1 (Figure 3-1). 数据管理周期中的人工数据审阅和编辑



3.1 集中审阅问卷及表格

调查问卷和表格的集中审阅，以及人工数据审阅，都需在数据输入之前做好，这样才能使每一项都有一个明确的、非缺失值可以键入。这样，数据输入人员就不需要自己去想该输入什么值（从而引入错误）。

对于 CFCS 项目，我们建议，人工数据审阅按调查工具来进行，而不是按个案（“家庭

组”或者“母子队”）来处理。例如，可以一次审阅多份**身体活动问卷**，而不用把所有属于同一个案的问卷都先审阅完。这样一来，从审阅一份**身体活动问卷**中发现的问题，可以立即告知其余**身体活动问卷**的审阅。

审阅已完成的调查问卷时，请检查以下问题：

- 是不是所有的问题都有答案吗？
- 答案是否相互矛盾？（譬如上一题回答失业，但下一题填写工资）
- 是否所有问题的变量赋值的编码都是正确的？
- 正确遵循跳转指示吗？
- 在须填选“是/否”的表格里，每一行是否填写了一个回答?（即，检查是否只有“是”的那一行才被填写）
- 所有的需要填数字填的空格是否填上了数字？
- 对“其他”的回答是否合适？
- 某些问题是否有很高的无回答率（缺省值）？
- 是否有调查员面试时写得笔记，可以用来帮助解决数据不一致的问题？
- 是否有需要跟踪解决的错误？
- 是否有其他的问题，需要跟踪解决？ （例如，难以辨认的字）

此外，在人工检查已完成的调查问卷时，可能会注意到一些不太对劲的规律，这可能会揭示出调查员的潜在错误。如表 3-1 (Table 3-1) 所示，所有的回答都是“有”或“是”，这要标明注意。我们可以理解，在任务 1 预试验的问卷中，有一些问题的回答如此，但不太可能所有的回答都是这样。有可能母亲并不全懂这些题或者调查员圈错了。此外，虽然不加以规定，我们当期许调查员发现回答都是“有”或“是”的时候，会做些笔记， 写下父母为什么这样回答的具体原因。有这种情况时，审阅人员应标明此调查问卷应进行编辑。

表 3-1 (Table 3-1). CFCS 任务 1 预试验问题（范例）

No.	Question	Answer
	Developmental Milestones	
M171	与其他同年龄的孩子相比，您的孩子是否在开始学坐、站立或走路方面比别人明显滞后**(**延迟、缓慢**)?** 是…… 1 否…… 2 备注: _____	
	Vision	
M172	您的孩子视力有没有问题**?** 有…… 1 没有…… 2 备注: _____	
	Hearing	
M173	您的孩子听力有没有问题**?** 有…… 1 没有…… 2 备注: _____	
	Comprehension	
M174	您的孩子在理解别人的语言方面有没有问题**?** 有…… 1 没有…… 2 备注: _____	

进行这种级别的人工审阅，最好是由充分理解调查工具中的每项问题的用意和使用的专业人员集中来进行。例如，营养师是人工审阅膳食日记和食物频率问卷的最权威的专业人

员，因为营养师可以体察到营养物质编码的细微差别。在美国，营养师使用核对表进行审阅。在附录中，我们提供有 CFCS 项目的审阅表（附录 A，B，C）。表 3-2 (Table 3-2)简介每个核对表的效用。详细指示在审阅表里。

表 3-2 (Table 3-2). 膳食日记和食物频率问卷 (FFQ) 的人工审阅核对表用途

名称	用途
附录 A. 膳食日记核对表（每日）	评估和记录调查员完成的食物日记的质量
附录 B. 膳食日记记录表（每日）	对受试对象的报告及调查员审阅的膳食日记进行质量评估，并将其量化。膳食日记记录表显示了受试对象是否正确地完成了膳食日记，调查员添加了所需的详细内容。
附录 C. 食物频率问卷（FFQ）核对表	评估和记录此调查员管理的食物频率问卷（FFQ）的质量

3.2 人工数据编辑

在人工审阅时，数据管理人员，要确保每个调查的问题，都有明确的数值或字符以供键入。管理人员依照编辑规范人工检查每个回答，并确定哪些问题是要编辑的，以及所需编辑的类型。为了保持格式的一致性，编辑规范应规定字符数和所用代码的范围，例如，日期的数据先列出年还是月呢？应为 YYYY/MM/DD（2012/10/25）还是 MM/DD/YYYY（10/25/2012）呢?

有时候，受试对象或调查员圈选了多重回答，但是问卷问题本身只可选一个回答。如果无法询问调查员哪一个才是正确回答，有些方式可以去除不恰当的多重回答，但也有可能会带来偏见。所以，事先应与项目负责人讨论要遵循的规则，将编辑多重回答的规范写下来。以下提供实例。

图 3-2(Figure 3-2)中的问题，其答案有一定的自然次序。它是问受试对象的教育程度，此题只取一个回答但两个答案被标选。也许这个受试对象完成了高中学业，并上了短短两年的大学。因此，高中和大学都打了标记。有时在美国，研究人员或许会用这样的方法做决定：（1）掷一枚硬币，如果 “头”朝上，就选择较高的学历（大学），“底“朝上就选择较低的学历（高中）；或（2）制定一条规则，总是选择较为保守的答案，这样，就要选择较低的学历（高中）。第一种方法引入了随机误差，而第二种方法则会引入系统性偏见。

图 3-2 (Figure 3-2). 关于教育程度的矛盾答案

104. 你的受到的最高教育是什么?

- [1] 小学和中学
- [2] 高中
- [3] 大专/大学

图 3-3 提供了一个口腔健康问卷调查的例子。这个问题也有一个自然的秩序。受试对象圈选了 1 和 2，分别代表“非常频繁”和“经常”。随机掷硬币或选择较温和的反应（譬如：经常），这两个方法之一都可被项目决定采用。但是，如果有人圈选了彼此不相邻的两个答案，比如“非常频繁” 和“偶尔”，则没有办法在人工审阅和编辑的过程中，有理可循地解决这个差异问题。

图 3-3 (Figure 3-3). 关于经常性的矛盾答案

2. 在过去的 12 个月内，在你的嘴里的任何地方，你有过很痛苦的疼痛吗？你会说...

1. 非常频繁
2. 经常
3. 偶尔
4. 几乎没有
5. 从来没有

图 3-4(Figure 3-4)是一个没有自然顺序回答的问题，所以之前所讨论的方法不适用。如果不与受试对象直接交流，了解最确切的答案，你就无法在人工编辑审阅过程中编辑这个回答。这个例子中，受试对象圈选了多个答案（“别人家”和“自己家”）。受试对象可能有时出差，所以，有时在自己的公寓住，有时在朋友家中住一些日子。因为没有自然顺序可循，无法在人工审阅和编辑的过程中解决差异。

图 3-4 (Figure 3-4). 关于居住问题的矛盾回答

H4HR2 你现在住在哪里？即，你经常在哪住的时间比较长？

1. 你的父母家
2. 别人家
3. 自己家（公寓，独立房屋，移动式房屋 等）
4. 集体宿舍（宿舍，军营，团体之家等）
5. 无家可归者，即，你有没有固定的地方住宿
6. 其他

当数据管理人员无法解决差异问题时，最佳的办法是由调查员做数据追踪，实地询问受试对象来收集一个回答。令一种方法则是由数据管理人员会插入代码来表示多重回答，无法解决。如此，最后的数据库内也可以查到。这类代码叫做 “一致性代码”，也用于 “不知道”， “拒绝”等回答，或像之前的例子标明是个“问题数据”。当人工数据编辑时结论无法解决差异问题时，数据管理人员会插入代码。分配给这些代码的值，应先征得项目负责人的同意。负值经常会被用来作为这些代码的值，因为它们不太可能与实际数据相混淆。一般来说，尽可能不要使用空格和零作为代码来表示缺失的数据，因为空格和零往往会与实际数据相混淆。表 3-3(Table 3-3)显示了在 RTI 通常使用的一致性代码。变量宽度视变量赋值而定加上负值符号（-）的宽度。譬如，年份的宽度是 4 个输入键（就像：2012），那末“不知道”的代码就该是-4444。

表 3-3 (Table 3-3). RTI 常用的一致性代码

代码	标签	形容	注解
-1	从不知道	此回答受试对象从不知道，现在和将来也永远不会知道，所以不要尝试更新此数据	特殊情况才用
-2	难以辨认	无法确认的回答	
-3	不适用	不适用	
-4	不知道	数据收集人笔记标明这个问题受试对象不知道答案。	常用的代码
-5	超过允许值范围的回答	回答超过规定字数或允许值范围内（例如，不能有月=13）	
-6	多重答案	对只要求有一个答案的问题，受试对象给了多个答案，而这些多个答案无法合理解释或解决。	
-7	拒绝	受试对象通过书面声明或在接受采访时拒绝回答	常用的代码
-8	空缺或无答案	受试对象应该回答却没有回答	
-9	合理无答案	受试对象不应该回答并且没有回答的问题（例如，按照规定跳过此题）	

数据管理人员应使用编辑问题表（见表 3-4）来记录问题，包括澄清的不完全清楚的回答，辨识很难读懂的字体，解决或消除旁注，消除不适当的一题多答等。根据此编辑问题表，数据管理人员可以联系调查人员，以确认手写的原始数据，参阅笔记，甚至到现场再进行数据搜取。对 CFCS 项目来说，能保存每个母亲和孩子组成的母子对（或用 ID 保存的母子对链接）的编辑问题表，可作为保证数据质量的重要文件，并可用来设计以后的调查员培训内容。此外，编辑问题表的内容也可以作为数据追踪收集的依据。

表 3-4 (Table 3-4). 编辑问题记录表

编辑问题记录表 项目: CFCS 任务 1 Fa ID 号.: 123456789 组号.:					
工具	问题 #	问题记录	解决办法	解决人:	日期
母亲	M005	检查工作经历 1988 - ?			
	M076, M077	检查哪一个应该标“是/有”			
母子	M131	问母亲这个答案是否应该是 1 或 3 ， 如果孩子不在学校时和父亲或母亲住，这个答案应该是 1 。			
父亲	F022	检查酒精含量 00.1 或 01.0?			
孩子	C019	每瓶检查酒精含量 00.1 或 01.0?			
母亲	M005	检查工作经历 1988 - ?			
母亲	M076, M077	检查哪一个应该标“是/有”			

图 3-5(Figure 3-5)是一个经过人工数据审阅和编辑的简单调查问卷。在问题的左侧是每项的数字代码， 这个数字代码是在人工数据审阅和编辑之后给出的。在人工数据审阅和编辑过程中， 数据管理工作人员，依照从编码本/数据字典开发出来的编辑规范，审阅每个答案，并为其赋值。根据表 3-5(Table 3-5)所示的编码本，答案有两个值，代码 1 表示“是”，代码 2 表示“否”。编辑如下：

- 问题 1 的回答为“是”（构选了“是”），被分配代码 1 表示“是”。
- 问题 2 被分配到代码 2，因为答案的类别标记是“否”。
- 问题 3，什么也没有标明。此对应的代码是“-8”，这是“空白或无答案” 的预定的一致性代码（见表 3-3, Table 3-3）。
- 在第 6 题，受试对象同时选择了 “是”和“否” 两个答案，代码是“-6”。在一致性代码表上（表 3-3， Table 3-3），“-6”意思是“多重回答”。

图 3-5 (Figure 3-5). 人工数据审阅和编辑的调查问卷

人工编辑	问题	是	否
1	1. 你是否失聪（听不见），或有严重的听力障碍吗？	☑	☐
2	2. 你是否失明，或即使戴着眼镜也很难看到东西吗？	☐	☑
-8	3. 由于身体，精神或情绪的不良状态，你是否注意力完全不能集中，不能记忆，或不能作出决定吗？	☐	☐
2	4. 你走路或爬楼梯非常困难吗？	☐	☑
2	5. 你穿衣或洗澡是否有困难？	☐	☑
-6	6. 由于身体，精神或情绪状态，你能不能独自出门，比如能不能独自去看医生或购物？	☑	☑

表 3-5 (Table 3-5). 用于图 3-5 的问卷编码本

变量名称	标签	类型	字符宽	值	变量标签
P2DAHEAR	失聪	数字型	2	1 2	是 否
P2DASEE	失明	数字型	2	1 2	是 否
P2DACON	注意力	数字型	2	1 2	是 否
P2DAWALK	走路	数字型	2	1 2	是 否
P2DADRES	穿衣	数字型	2	1 2	是 否
P2DAERR	琐事	数字型	2	1 2	是 否

练习 3-2　你会如何编辑这个CFCS问卷？

使用以下编辑规范来编辑五行身体活动问卷（PAQ）的问题答案。　（提示：可用表 3-5/Table 3-5 中的一致性代码）

变量名称	变量值宽度	范围	描述
(A)是否参加	2	1，2	如果变量“是否参加”=02，以下变量不回答：为几个月；每月几天；每周几天；多长时间
(B)几个月	3	1-12	总数不能是 0；如果是个位数，前面要加 0 （如 01，02）
(C)每月几天	3	1-31	总数不能是 0；如果是个位数，前面要加 0 （如 01，02）
(C)每周几天	2	1-7	总数不能是 0；如果是个位数，前面要加 0 （如 01，02）
(D)多长时间	3	HH	如果小时是个位数，，前面要加 0 （如 01，02）
(D)多长时间	3	HH	如果分钟是个位数，，前面要加 0 （如 01，02）

	(A)是否参加 1 是 2 否	(B)一年中有几个月进行该活动？	(C)在这些月份中，平均每月（或每周）有几天进行该活动？ |___|___|天/月或/|___|天/周	(D)这些天中平均每天进行多长时间？ |小|时|:|分|钟|
坐着/躺着：看电视、录像或影碟	|1|	|___|0|	|___|___|/|___|	|___|___|:|___|___|
坐着/躺着：看电视、录像或影碟	|1|	|1|2|	|3|0|/|7|	|_|:|3|0|
坐着/躺着：看电视、录像或影碟	|1|	|_|2|	|_|/|5|	|6|_|:|3|0|
坐着/躺着：看电视、录像或影碟	|2|	|_|_|	|_|/|_|	|_|:|_|_|
坐着/躺着：看电视、录像或影碟	|2|	|0|0|	|0|0|/|0|	|0|0|:|0|0|

答案：

	(A)是否参加 1 是 2 否	(B)一年中有几个月进行该活动？	(C)在这些月份中，平均每月（或每周）有几天进行该活动？ |___|___|天/月或/|___|天/周	(D)这些天中平均每天进行多长时间？ |小|时|:|分|钟|
坐着/躺着：看电视、录像或影碟	|~~1~~| *2*	|__|~~0~~| *-99*	|___|___|/|___| *-99 / -9*	|___|___|:|___|___| *-99 : 99*
坐着/躺着：看电视、录像或影碟	|1|	|1|2|	|3|0|/|7|	|*0*|*0*|:|3|0|
坐着/躺着：看电视、录像或影碟	|1|	|*0*|2|	|___|___|/|5| *-9 9 /*	|~~6~~|___|:|3|0| *0 6*
坐着/躺着：看电视、录像或影碟	|2|	|~~0~~|~~0~~| *- 9 9*	|___|___|/|___| *-9 9 / -9*	|___|___|:|___|___| *-99 : 99*
坐着/躺着：看电视、录像或影碟	|2|	|~~0~~|~~0~~| *-9 9*	~~|0|0|/|0|~~ *-9 9 / -9*	~~|0|0|:|0|0|~~ *-99 : 99*

3.3 创建编码本或数据字典

编码本（数据字典）是一个列出数据文档的内容，并描述其特点的文件。简单地说，它记录了变量的规则。一旦数据需被人工审阅及编辑，就必须先创建一个编码本，以用于编写数据的输入规范（给数据输入人员用的规范）。

一个的编码本应简单如实地描述数据。它按照每个变量出现的先后顺序给它们命名，描述变量的格式，并列出它可能拥有的值，包括空白，拒绝，不知道等的缺值代码。数据的格式，无论是数字（比如，用于计算平均值）、文字（用于分类或计数）、或日期，都应该对最终使用这些数据的分析师是有意义的。首先，要与项目主要负责人和其他需要使用的数据的人讨论编码本的内容，这是至关重要的。因为这个讨论的决定，将影响在最终的数据集里的数据的式样。

在某些情况下，你可以用原有问卷来创建一个基本的编码本。将问题改为变量名。试想一下，有一个问题是这样的：

1. 上周，你有没有吃任何在快餐店购买的食物？

你可以将变量命名为 Q1(指问题 1），或命名为 EASTFAST(快餐,根据问题的内容命名），或命名为一个组合词，如 Q1EATFAST(题 1 快餐)。要使用你和你的项目合作者都可以理解的命名系统，并且该系统对调查问卷上所有的问题都保持一致。接下来，提供一个变量的标签，如“上周的快餐“。如果问题很短，你可以直接使用该问题的用词作为变量的名称。读一下变量的名称和标签，数据分析师就可以知道如何引用这个问题和这个问题是在问什么。接下来，分析师会想知道是一个什么样的变量值。一个受访者，在回答这个有关快餐的问题时，只要圈“是”或“否”。有时可能会遇到没有答案的问题或空白。编码本应该标明对以下各种情况，数据输入人员应该键入一个什么值：1（是）， 2（否），还可能需要-8（无答案）。

这是编码本对这个问题的编码（它也可以是表格的形式）：

Q1EATFAST（上周在快餐店），数字。

1. 上周，你有没有吃在快餐店购买的任何食物？

1 是
2 否
-8 无答案

一些计算机程序无法读取超过 8 个字符的变量名。统计程序只打印前 20 个字符的变量标签。在创建编码本之前，一定要与您的协作者和数据用户（如果可能的话）确定一下，应该使用多少个字符的变量名，变量标签可以使用得最多的字符数是多少。

练习 3-2. 创建一个有很多变量的问题的编码本

下面的表格中的 Q2 是有关昨天餐饮的问题，这个问题允许受访人圈选两个或两个以上的答案。每个圈选的答案意味着“吃了一顿饭”，用代码 1 表示。每个未圈选的答案意味着“没吃这顿饭”，用代码 2 表示。

a. Q2 将产生多少个变量？

b. 如果整个 Q2 被跳过时，变量应该用什么样的代码？ （提示：请参阅表 3-3/ Table 3-3）

c. 编码本应该是什么样子的？

适合老年人的饮食问卷
Q1 上周，你有没有吃从快餐店购买任何食物？ 是 否
Q2 你昨天吃了哪些餐？ （选择所有适用的） A）早餐 B）上午点心 C）午餐 D）下午点心 E）晚餐 F）晚间点心 G）其他 描述其他餐：_____

研讨练习 3-2 的答案

这个问题，Q2 将一共产生八个变量。前七个变量可被命名为 Q2a，Q2b，Q2g 或者 Q2Break，Q2AMSnk，Q2Other， 第八个变量，则是受访者用自己的话来形容“其他”饭的文字回答。

您将创建变量的描述性的标签，如“昨天吃了早餐”，“昨天吃了上午点心”等。每个变量，从 Q2A Q2F，将采取 1 或 2 的值。根据你创建的一致性代码，如果受访者跳过整个这个问题，你可以将这七个变量都输入-9。

对于变量 Q2H（文字符），没有办法指定此文字变量的值和标签，因为内容可根据受访者的经验而有所不同，是不可预知的。这不同于对有特定的数字或字符，可以作出有含义的编码。数据输入规范应规定这种变量的最多字符数，比如可以是 10 或者 20。

编码本：适合老年人的饮食问卷					
变量名称	变量标签	变量类型	变量宽度	变量赋值	赋值标签
Q1	上个星期吃了快餐	数字型	2	1	是
				2	否
				-9	空白
Q2a	昨天吃了早餐	数字型	2	1	是
				2	否
				-9	空白
Q2b	昨天吃了上午点心	数字型	2	1	是
				2	否
				-9	空白
Q2c	昨天吃了午餐	数字型	2	1	是
				2	否
				-9	空白
Q2d	昨天吃了下午点心	数字型	2	1	是
				2	否
				-9	空白
Q2e	昨天吃了晚餐	数字型	2	1	是
				2	否
				-9	空白
Q2f	昨天吃了晚间点心	数字型	2	1	是
				2	否
				-9	空白
Q2g	昨天吃了其他餐	数字型	2	1	是
				2	否
				-9	空白
Q2h	描述其他餐	字符型	20		

此外，一些研究使用日记。三天紫外线日照记录就是一个很好的例子。（见图 3-6; Table 3-6）受访人在他或她的表格的行和列中填写答案。这种类型的表格包含了大量的信息，但编码本应遵循前面所描述的同样的原则。

图 3-6（Table 3-6）. 三天紫外线日照记录 （范例）

日期： 2012 年 03 月 01 日

今天是否佩戴紫外线徽章？ ☑ 是 ☐ 否

徽章编号： 888888

今天的佩戴有无出现问题（例如有没有带满全天，有没有弄脏、弄湿，有无破损等）？ ☑ 无 ☐ 有（请注明）：

	户外日光暴露时间（在对应时间段内方框内打勾☑）					是否使用过防晒霜	如佩戴或使用，在方框内打勾			衣着类型（根据“衣着指南”，圈选相应的数字）			
	0 分	<15 分	15-29 分	30-44 分	45-60 分	如使用，在方框内打勾	手套	太阳镜	遮阳伞	帽子	上身	下身	鞋
上午													
5-6 点前	☑	☐	☐	☐	☐	☐	☐	☐	☐	① 2 3	① 2 3	① 2 3	① 2 3
6-7 点前	☐	☐	☐	☑	☐	☐	☑	☐	☐	1 ② 3	1 ② 3	① 2 3	1 2 ③
7-8 点前	☐	☐	☑	☐	☐	☑	☐	☐	☐	① 2 3	1 2 ③	1 2 ③	1 2 ③
8-9 点前	☐	☑	☐	☐	☐	☐	☐	☐	☐	① 2 3	1 2 ③	1 2 ③	1 2 ③
9-10 点前	☑	☐	☐	☐	☐	☐	☐	☐	☐	① 2 3	1 2 ③	1 2 ③	1 2 ③
10-11 点前	☑	☐	☐	☐	☐	☐	☐	☐	☐	① 2 3	1 2 ③	1 2 ③	1 2 ③
11-12 点前	☑	☐	☐	☐	☐	☐	☐	☐	☐	① 2 3	1 2 ③	1 2 ③	1 2 ③
下午													
12-1 点前	☐	☑	☐	☐	☐	☑	☐	☐	☐	① 2 3	1 2 ③	1 2 ③	1 2 ③
1-2 点前	☐	☑	☐	☐	☐	☐	☐	☐	☐	① 2 3	1 2 ③	1 2 ③	1 2 ③
2-3 点前	☑	☐	☐	☐	☐	☐	☐	☐	☐	① 2 3	1 2 ③	1 2 ③	1 2 ③
3-4 点前	☑	☐	☐	☐	☐	☐	☐	☐	☐	① 2 3	1 2 ③	1 2 ③	1 2 ③
4-5 点前	☑	☐	☐	☐	☐	☐	☐	☐	☐	① 2 3	1 2 ③	1 2 ③	1 2 ③
5-6 点前	☑	☐	☐	☐	☐	☐	☐	☐	☐	① 2 3	① 2 3	① 2 3	① 2 3
6-7 点前	☑	☐	☐	☐	☐	☐	☐	☐	☐	① 2 3	① 2 3	① 2 3	① 2 3

数据输入人员的数据输入规范应表明，数据是从左至右，从上到下输入的。编码本（见图 3-7; Table 3-7）按照这个的顺序描述各个变量，从第一个变量—“日期”到最后一个变量—“在晚 7 点到 8 点之间是否用了遮阳伞”。

图 3-7 (Table 3-7). 三天紫外线日照记录编码本

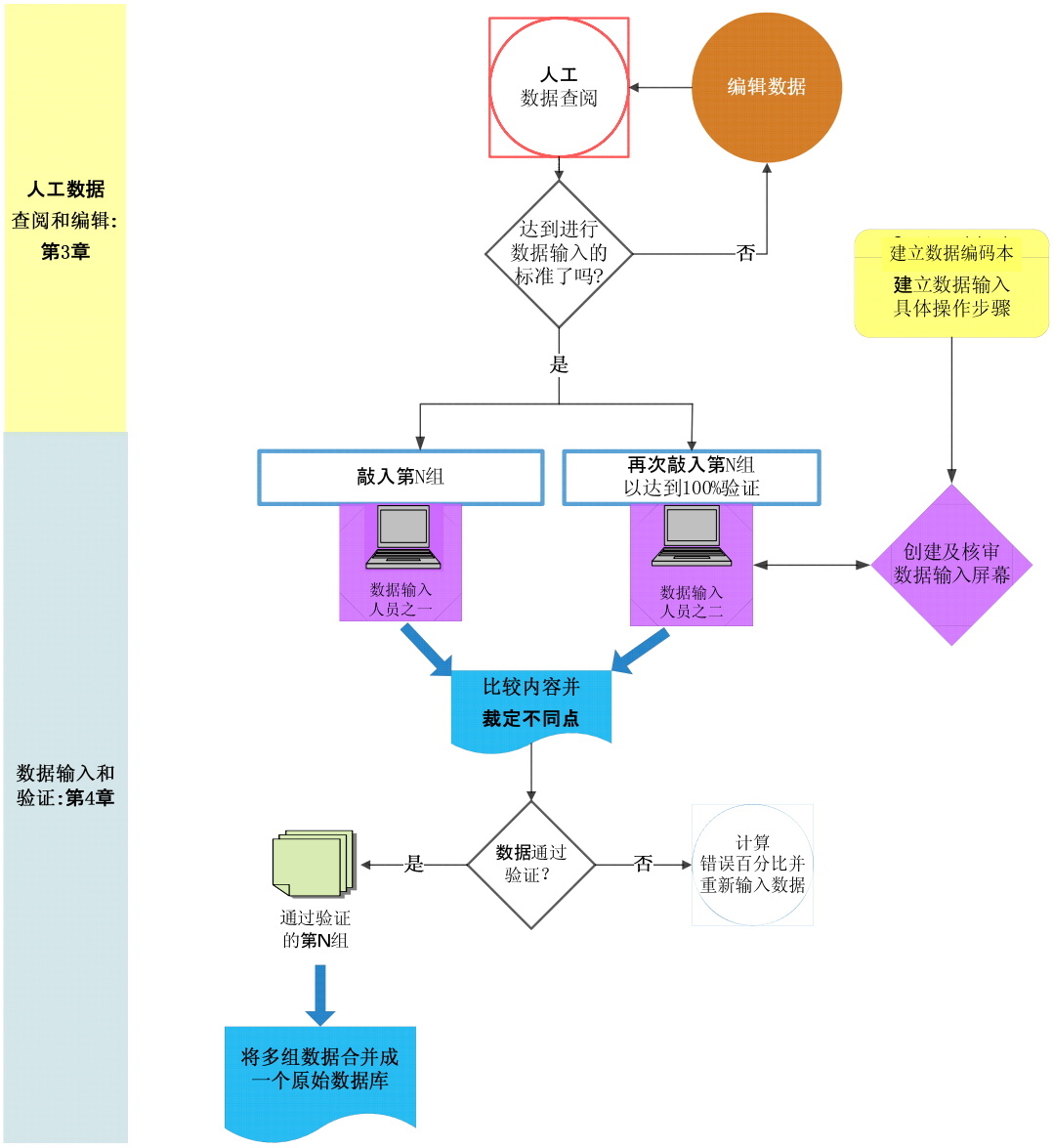
变量名称	变量标签	变量类型	变量宽度	变量赋值	赋值标签
date_1	Day 1_Date	日期	6		
ID	Participant ID	数字型			
badge_1	Day 1_Did you wear badge?	数字型	1	1 2	是 否
badgeID_1	Day 1_Badge ID	数字型	6		
ifProblem_1	Day 1_Did you have problems with badge?	数字型	1	1 2	有 无
problem_1	Day 1_Please specify the problem	字符型			
out_A_1	Day 1_Outdoor time at 5-6 AM*	数字型	1	1 2 3 4 5	0 <15 15-30 30-45 45-60
sunScreen_1	Day 1_Did you apply sunscreen today*	数字型	1	1 2	是 否
SC_01_1	Day 1_Body site for sunscreen use_site 1*	数字型	1	1 2	是 否
GL_A_1	Day 1_If wore gloves at 5-6 AM*	数字型	1	1 2	是 否
SG_A_1	Day 1_If wore sun glasses at 5-6 AM*	数字型	1	1 2	是 否
UM_A_1	Day 1_If used umbrella at 5-6 AM*	数字型	1	1 2	是 否
UPPE_A_1	Day 1_Clothing code for upper body at 5-6 AM*	数字型	1	1 2 3	
LOWE_A_1	Day 1_Clothing code for lower body at 5-6 AM*	数字型	1	1 2 3	
HEAD_A_1	Day 1_Clothing code for head wear at 5-6 AM*	数字型	1	1 2 3	
FOOT_A_1	Day 1_Clothing code for foot wear at 5-6 AM*	数字型	1	1 2 3	
Continued in similar fashion for the next 14 rows of diary, through 7-8 PM time period.					

4. 数据输入和验证

数据输入和验证过程的目的，是从文件资料中提取数据，并验证数据输入的准确性。从所有调研工具中得到的数据，都应键入的数据程序，包括调查问卷，笔记，测试结果或测试表格。输入的数据，包括遗漏数据的一致性代码，要根据第 3 章中所描述的步骤进行的数据编辑。

图 4-1(Figure 4-1)是根据数据输入规范，建立的和测试过的数据输入屏幕。如果数据被分成组，数据输入人员则以每一个组为单位，打入和核实输入的数据。每组被核实的数据被编入一个原始数据库，以进一步进行数据清理。

图 4-1 (Figure 4-1). 在数据管理过程中的数据输入和验证的流程



4.1 数据输入

在第 3 章中所讨论的，数据输入的规范是根据密码本研发出来的。Blaise，Microsoft Access，和 EpiInfo 都是可用来编程数据输入屏幕的流行数据系统。一旦数据输入屏幕研发出来，就应该对它们进行测试，以确保它们遵循规范。

对调查工具的编辑和编码，必须在人工审阅过程中做，在数据输入之前做好。正如在前面第 3 章所讨论的，在人工审阅的过程中，工作人员必须确保每个调查的题目都包含明确的可键入的数值或文字。这样，数据输入人员则可以看到什么就输入什么，并将任何不清楚的问题交给数据管理人员处理。数据输入人员不需要假设任何事情。数据输入系统中的任何变量的输入，都不应该是空白。如有数据遗漏，则需按照如同第 3 章表 3-3 (Table 3-3)的输入指定代码。

数据的格式，决定了数据如何被键入。有三种最常见的格式：代码，数字型和字符型。

- “选一个”（见图 4-2，Figure 4-2），调研对象必须选择一个答案,而与这个答案相对应的代码（数字）将被键入。

图 4-2 (Figure 4-2). 用选择数字来回答问题

M151.是谁首次告知您的孩子患有肿瘤？

村医 1
乡镇或区卫生院医生 2
县或市医院医生 3
省级医院医生 4
肿瘤医院医生 5
儿童医院医生 6
其他，请说明 _____ 7

- 数值（参见图 4-3; Figure 4-3），受试对象给与数值，此数值将被键入。

图 4-3 (Figure 4-3). 通过填写数字来回答问题

M150. 在您第一次被告知孩子患有肿瘤时，孩子多大? [周岁年龄和月龄]
如小于1 岁，年龄填“00”，填写实际月龄

年龄(如不知道填“99”) ..... __ 11__ 岁
月龄 (如不知道填“99”) ..... __ __ 月

输入填写的数字答案

- 字符型的开放式答案（见图 4-4; Figure 4-4），受试对象提供一个开放式（自己写的）文字答案,而这个答案将被作为文字[字符型数据]，输入到数据输入程序里。

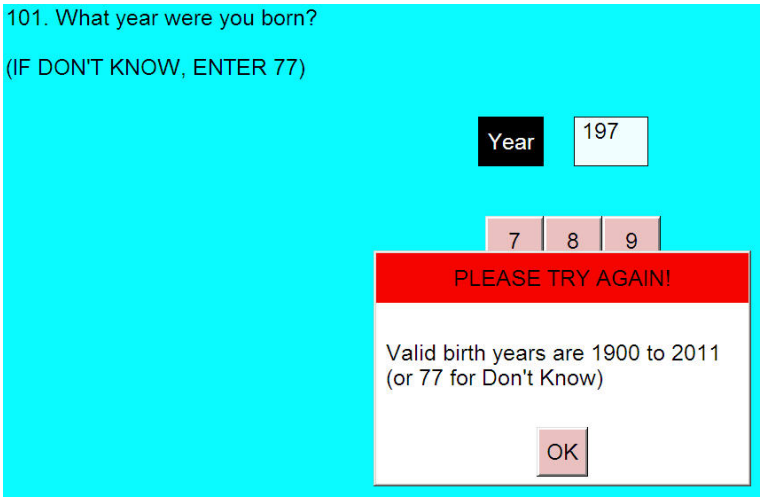
图 4-4 (Figure 4-4). 通过书面文字回答问题。

M152. 您的孩子首次被确诊为肿瘤的医院的名称和地址:
中国河北唐山市老亭区金融街老亭医院

将此开放式答案，逐字逐句地输入

范围检查很重要。范围检查， 将输入数据程序的输入值和该变量可接受的范围值相比较。当数据输入人员输入的值超出了可接受的范围，数据程序将出现错误信息警告，警告输入人员这里出现了问题。使用 EpiInfo 的数据输入系统为例，图 4-5(Figure 4-5)显示了一个错误信息警告，输入的出生年份是在可接受的年份范围之外。该系统将不允许数据输入人员输入下一个变量，直到此错误数字被修正。如果使用其他数据输入系统，也应通过编程，来做相似的范围检查。

图 4-5 (Figure 4-5). EpiInfo 的范围检查示例



再试一次！

从 1900 年到 2011 年之间才是出生年可接受的范围是。

练习 4-1. 什么是此项CFCS数据的检查范围？

这里是一道 CFCS 的问卷题，其标准数据答案是要写入编码本里的。你会如何设定数据的检查范围？（提示：可接受的范围根据可键入的数值来定。）

Q1EATFAST（上周快餐）数字输入。

1。 上周，你有没有吃任何从快餐店购买的食物？

1 有
2 没有
-8 无答案

4.2 数据验证

如第 2 章（接收记录及分组流程）所述，将实体文档资料(如纸质文件)分组（如第 1 组，第 2 组等），这样就更容易追踪文件目前正在哪里,以及他们在数据管理过程的现状。有时，数据管理人员会决定将所有表格和问卷放在同一组。分组系统之后，第一组的数据被输入到数据输入系统。只要有可能，我们推荐 100％的数据验证。这意味着这组数据一定会被重新键入。

验证级别可以包括单次重新键入（对同一变量创建两个键入的记录）或双次重新键入（对同一变量创建三个键入的记录）。对同一组数据，可以由不同的数据输入人员作重新键入。但是，第二次键入数据的数据输入人员,一定不可以是第一次键入数据的人。数据输入可以双盲进行（即，输入员 2 看不见输入员 1 输入的数据）。该系统也可设计成这样的方式，当第二个输入员键入数据时，该数据与第一次输入的原始数据同时实时进行比较验证。

如果表格的数量很少，所有数据的准确性则更是至关重要。我们建议使用单次重新输入法，由不同的数据输入人员，独立输入数据两次。然后，将数据进行比较，看其是否一致。当发现差异，审阅人则需察看实体文档资料(如纸质文件)，以确定输入正确的数据。这个审阅人的角色常常可由第二个键入数据的人扮演。当数据质量需要进行密切监测时，也可能需要由一个不属于数据输入公司的，独立审阅人来审阅。若聘请数据输入公司来研发输入屏幕及键入数据，要确保其遵循本指南内的程序。项目主管（数据管理人员和领导）应该检阅数据输入的规范，且评估输入屏幕符合编码本（数据字典）。数据输入公司应提交所有两个键入的记录给中国疾病预防控制中心的数据管理人员审阅。

RTI 力图误差率小于 1％。根据数据量的大小，误差率也许会在过了一定时间后再作计算。每个人的错误率是根据输入变量的总数与错误的总数之比来计算的。例如，如果一个数据输入人员键入 1000 个变量，但出了 40 项错误，那么他的误差率是 4％（40/1000 =0.04），这是不符合可接受的标准的。如果有多个数据输入人员，就要计算总体误差率。例如，如果有 1000 个变量和甲乙两个数据输入人员共出了 40 项错误，则整体的误差率是 2％（40/1000/2= 0.02）。当某个数据输入人员的误差率大于或等于 1％，他/她就需要经过重新培训，以进一步理解数据输入的规范。然而，高误差率也许会揭示表格的数据输入规范或数据输入系统本身的可用性可能会有一些问题。

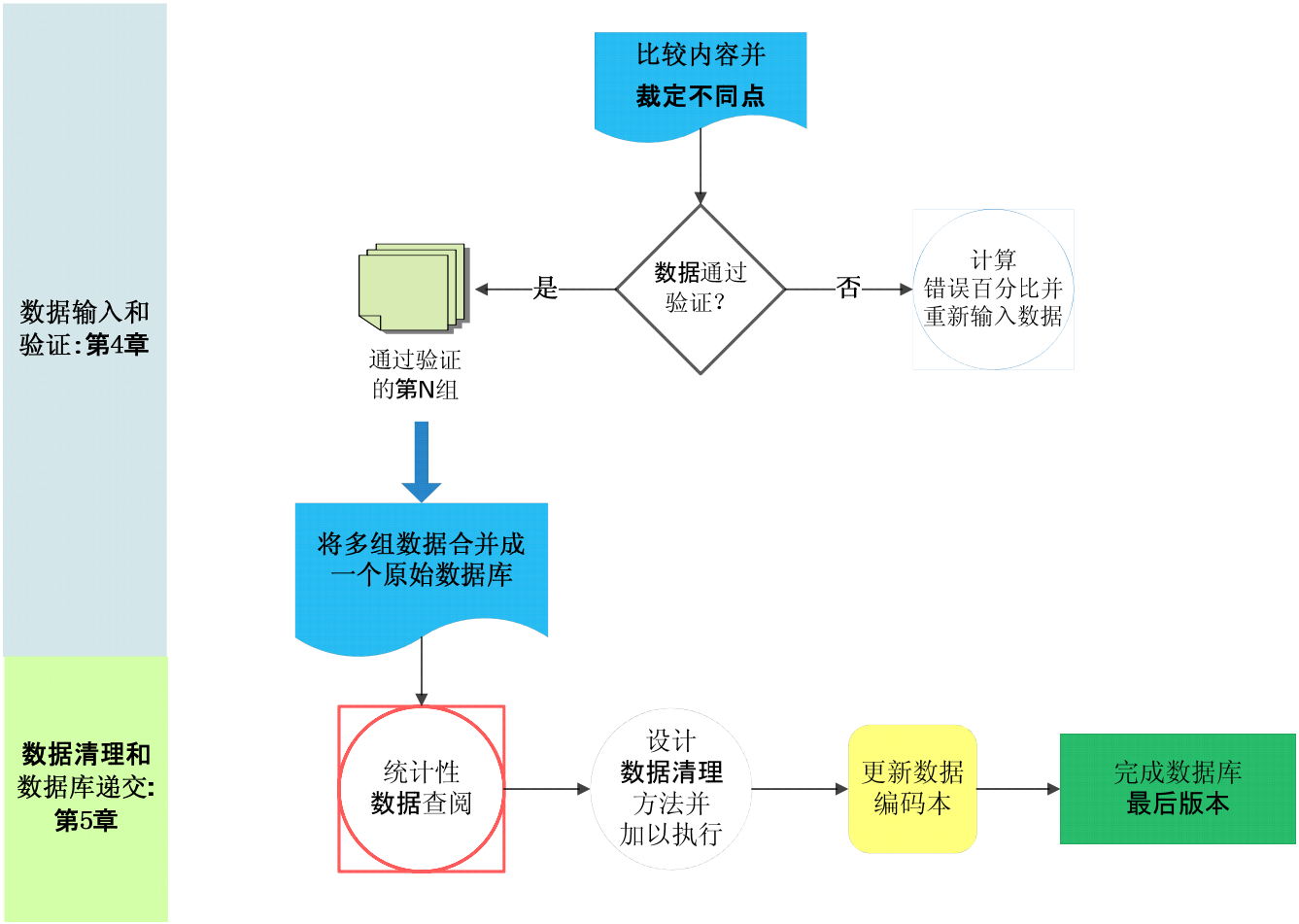
当所有的小组数据已被输入和验证后，将它们组合成一个单一的数据库，这样可以用的统计软件来对这个数据集进行数据清理。

5. 数据清理和数据库的递交

数据清理和数据库递交的目的是为了保证分析当案的完整性和准确性。这个过程是在数据输入完成并生成原始数据库之后才进行。

图 5-1(Figure 5-1)显示数据清理和数据库递交是整个数据管理过程的最后一步。

图 5.1 (Figure 5-1). 数据管理过程中的数据清理和数据库的递交



5.1 数据清理

虽然数据编辑,集中人工查阅,和数据输入验证都是用来最大程度减小失误的方法,这些步骤只能在单个问题或问卷的层面上进行。 而用计算机程序进行的对原始数据库的清理则能检测到聚合数据层面上的错误,以便编辑或删除数据库中重复,不完整,不一致或者不正确的数据。 这个步骤要分两步:

1. 用统计软件例如 SAS 来写数据检查的程序,生成数据输出结果用来评估
2. 使用数据结果来查找错误,然后人工对比数据和问卷来验证数据数值

请使用分组流程首页来查找实体文件材料,以找到完成的调查问卷,由此来判断是否有数据输入的错误。如果没有,研究项目的领导要对如何处理这些个案给予关注。

请阅读以下列出的一些基本的数据清理方法:

5.1.1 制作所有数据库内容的频率分布

查看频率分布是非常有用的,一般能从中发现很多错误。然后,能够追溯到错误的根源,帮助改善整个过程，也就是，错误是在数据输入时产生的吗? 我们是不是在人工数据查阅和编辑的步骤中忽略了这个错误?

数据符合常理吗?

有时仅通过看回答是不是符合常理就能发现错误。比如说,关于一个身高的问题,对于儿童来说,2 米的身体是不可能的。还有,查看数据是否有异常值，如太高或低的数值。 那些变量有空缺值?(有可能是问卷本身有问题或者是指示不清楚造成逻辑上的不连贯,但更大的可能是问题比较敏感或者不容易回答。)

有重复数据吗?

检查独特编码也就是个案数码的分布频率。在有些研究项目中,每个人只需填一个表。在这种情况下, 个案数码只应出现一次。但是,如果一个人需要填多个表, 此人的个案数码只出现了一次,这就说明有问题。独特个案数码有效吗,譬如其中的一个符号或数值不在本研究项目之内? 请看图 5-2 (Figure 5-2), 一个研究项目中,要求每个受访对象填一份问卷个案数码,但是却有两个数码为 0220838H. 另外还有一个后缀 “A”的个案数码，并不符合此项目数码范围，表示另一个样本框中的一位受访对象的数据被输入到了错误的数据库中。如果发生这类错误,要让调查人员以及数据输入人员知道。

图 5-2 (Figure 5-2). ID 数字样本的打印件,显示一个重复和错误的数据输入

```
proc freq data=renters.QC1;
tables caseid;
run;
```

The FREQ Procedure

Caseid

IDNUM	Frequency	Percent	Cumulative Frequency	Cumulative Percent
02002181	1	0.23	1	0.23
02006385	1	0.23	2	0.47
0204718H	1	0.23	3	0.70
02070384	1	0.23	4	0.93
02182383	1	0.23	5	1.17
02188383	1	0.23	6	1.40
02189383	1	0.23	7	1.63
0219038H	1	0.23	8	1.86
02202380	1	0.23	9	2.10
0220838H	2	0.46	11	2.56
02210382	1	0.23	12	2.80
02212387	1	0.23	13	3.03
02230383	1	0.23	14	3.26
0223938A	1	0.23	15	3.50
0224538H	1	0.23	16	3.73

格式和一致性代码是否一致?

请看整数或小数点的数字格式。在数据库中,它们的格式全都一致吗?例如,对于表示 30 分钟,数据或许会有以下几种格式,但应改为全都一致的格式。

00:30　　30　　0:30　　0030

另外,如第 3 章所说, 一致性代码的生成是为显示问题未回答或者逻辑不连贯的原因 (例如, 空白的答案, 拒绝回答, 不好的数据,错误的跳题)。

是否正确地按跳转题指示回答问题?

临近的问题便于评估。比如说,当案中有 100 条记录,代表着不同的问卷。其中,有 45 个问卷对一个问题的回答为“是”,而这些对于回答为“ 是”的问卷,后面都跟问 5 个必答的问题。也就是说,这 5 个问题都得有答案数据。如果有 1 个或更多的问题没有答案,这表明访谈对象没有回答这个问题,这个问卷也就是不完整的,或者数据输入出了问题。表 5-3(Table 5-3)显示那些回答吸烟的调查对象,被要求回答他们最开始吸烟的年龄。这说明调查对象正确地按照跳转问题要求的方式回答了接下来的问题。

图 5-3 (Figure 5-3). 用来评估跳转问题回答方式的调查分布频率

Have you ever smoked cigarettes (including hand-rolled or device-rolled)?

你曾经吸过烟吗(包括手卷烟或机器卷烟)?

Ever_ Smoked	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	45	45.00	45	45.00
No	46	46.00	91	91.00
Refused	7	7.00	98	98.00
DK	2	2.00	100	100.00

How old were you when you started to smoke?

你最开始吸烟的时候多大年纪?

Age_Smoking_ Start	Frequency	Percent	Cumulative Frequency	Cumulative Percent
12	2	4.44	2	4.44
14	1	2.22	3	6.67
15	2	4.44	5	11.11
16	8	17.78	13	28.89
17	9	20.00	22	48.89
18	23	51.11	45	100.00

Frequency Missing = 55

Frequency = 频率 Percent = 百分比 Cumulative Frequency = 累积频率
Cumulative Percent = 累积百比分

或者,也可以像图 5-4(Figure 5-4)那样将临近的数据值用竖列的形式打印出来。 这个图显示了 HOM2_60“否” 的答案跳过了下面的 5 个问题(HOM2_61, HOM2_62, HOM2_63, HOM2_64, HOM2_65),但是给出“是”的答案的访谈对象回答那 5 个问题。所有的访谈对象都回答了新的一部分(REFI)。最后一排(Obs 114)则说明表格没有全部完成。

图 5-4 (Figure 5-4). 数据以竖列形式打印出来的样本示例

```
proc freq data=renters.QC1;
tables hom2_60 hom2_61 hom2_62 hom2_63 hom2_64 hom2_65 refi;
run;
```

Obs	HOM2_60	HOM2_61	HOM2_62	HOM2_63	HOM2_64	HOM2_65	REFI
99	NO	.	.	.	.	.	1
100	NO	.	.	.	.	.	1
101	NO	.	.	.	.	.	2
102	NO	.	.	.	.	.	1
103	NO	.	.	.	.	.	2
104	NO	.	.	.	.	.	2
105	NO	.	.	.	.	.	2
106	NO	.	.	.	.	.	1
107	NO	.	.	.	.	.	1
108	NO	.	.	.	.	.	2
109	NO	.	.	.	.	.	1
110	YES	1	2	1	1	3	1
111	YES	2	1	2	1	1	1
112	YES	1	1	2	2	2	1
113	YES	2	2	1	1	2	2
114	YES	1	1	1	.	.	.

数据和数据编码本相符合吗?

按数据编码库里的顺序仔细看分布频率,里面会给出已定的数据的范围,变量名称,变量标签。图 5-5 (Figure 5-5)显示了一个实际数据和数据编码本中的变量名字和/或数据标签不相符的例子。结果显示数值“22”很明显不是应有的回答之一。

图 5-5 (Figure 5-5). 数据编码本与数据之间对于变量和变量标签的偏差

```
proc freq data=temp_QC1;
tables SCR18;
run;
```

Does the head of household own home?

SCR18	Frequency	Percent	Cumulative Frequency	Cumulative Percent
YES	38	34.23	38	34.23
NO	72	64.86	110	99.09
22	1	0.09	111	100.00

Frequency Missing = 4

数据编码本输入:

SCR18. 这个家庭户的户主拥有还是租住这个房子?

1 拥有
2 租住
-4 不知道
-7 拒绝回答

另外,数据中“竖列” 错位也会发生。这种错误的发生原因有几种可能,有时是在数据转换格式的过程中发生,有时是处理数据的人在复制和粘贴时出现失误,有时可能是数据输入的错误。 用对比数据和数据编码本的方法易于找出这种错误。

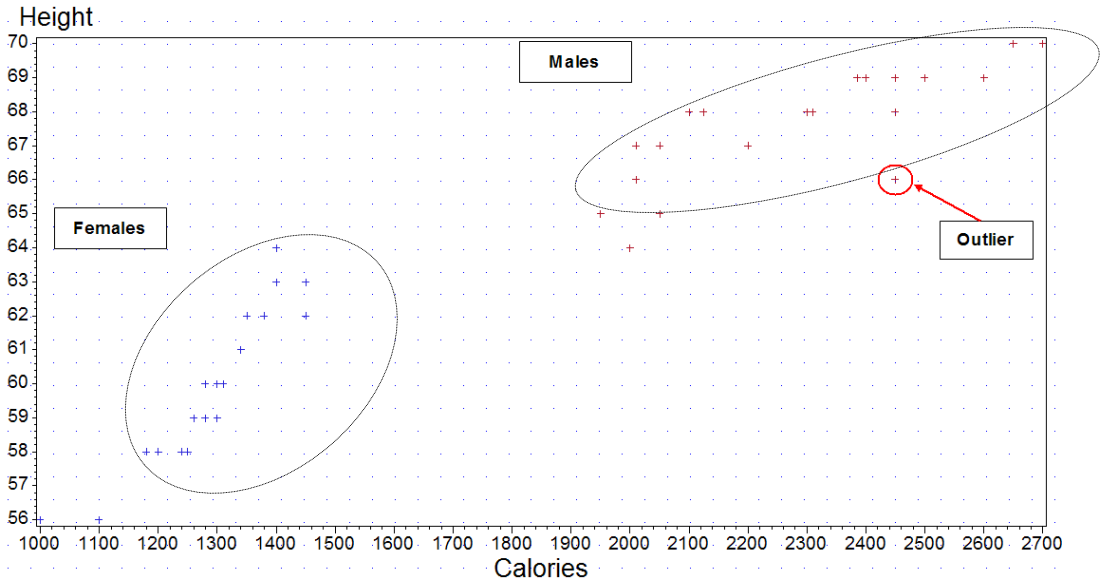
图 5-5 (Figure 5-5) 也是个竖列错位的例子,因为数据输出显示的是“是/否” ,而数据编码本则显示应该回答“拥有/租住。” 由于这种错误改变了数据,当在一个变量中发现时候,应该对其他的变量也同时查看一下。

当发现某受试对象的数据有偏差,用问卷的 ID 数码来找到这个问卷,判断是要更新数据编码本,还是要修正数据中的错误。数据编码本和实际的数据要保持一致。有时数据输入人员会因数值范围改变或需要更长大的文体输入而改变数据输入系统, 这样就需要数据管理人员更新数据编码本来反映这种情况。如果对数据编码本作出诸如此类的修改,一定要做好详细记录。 把新的数据编码本进行重新保存成另外的文件,但保有旧的档案名,或者用“跟踪显示修改内容”的功能, 或者用颜色来显示更改的内容。这样以后的文档资料就能看出系统是怎么逐渐改善的。文件的所有版本都要保留,需要有个版本控制表来列出何时何人创建了这些版本。这一做法也会对将来回答合作者和数据使用者对你递交的数据库提出的问题有所帮助。

5.1.2 使用散点图和流程图

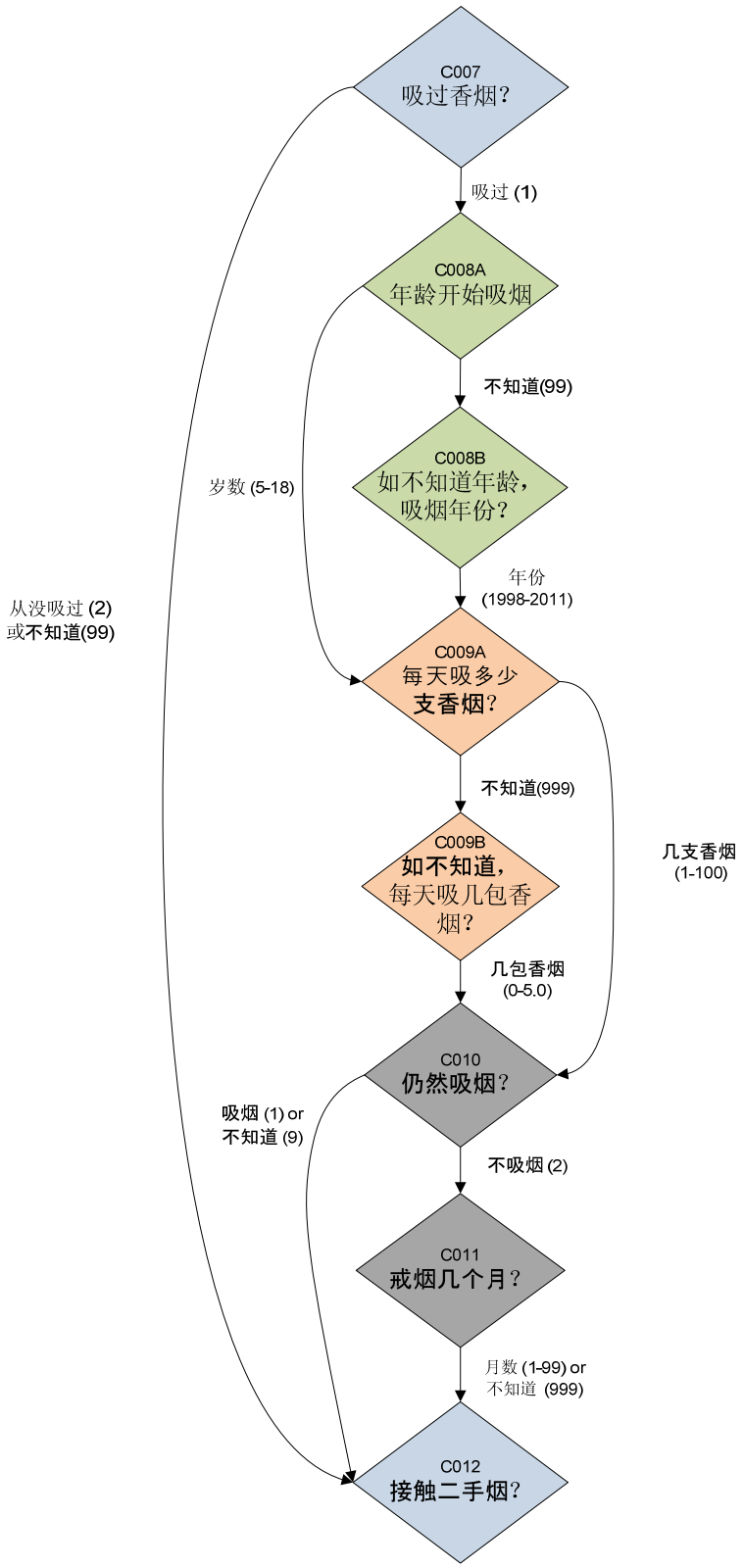
散点图有利于检验变量是如何相关的。对你认为可能相关的变量制作分布频率的散点图。图 5-6(Figure 5-6)显示男女身高和消耗的卡路里成正比。如果图形看起来不符合常理,要把结果向项目领导进行报告。你或许发现了数据编码弄反了的证据,或者是研究人群本身就有些不同寻常。

图 5-6 (Figure 5-6). 男女身高和日进卡路里的散点图



对有跳转问题格式的问卷,可画一个流程图。这样做会有助于更容易得看清楚访谈对象可能会按怎样的路线来回答问题。 图 5-7(Figure 5-7)显示了 CFCS 任务 1 预试验孩子问卷的一个例子, 关于吸烟的问题 C007-C012。 它显示了前一个问题的答案决定了调查对象随后被要求回答的问题。问题 C007 问调查对象是否曾吸过烟。如果答案为 “是”,调查对象要回答下个问题(C008A) 以及关于吸烟历史和行为的一系列问题。如果答案是从未吸过烟,或者调查对象主动说“不知道”, 那些问题就需要跳过,直接回答问题 C012。 下面几个问题询问在家里接受被动吸烟的天数。通过流览有复杂跳跃问题格式的流程图,你一眼就可以看出需要制作这样几个交叉制表: 问题 C007 和 C012, 问题 C007 和 C008A, 问题 C008A 和 C008B。

图 5-7 (Figure 5-7). CFCS 任务 1 预试验孩子问卷调查的流程图 (C007 - C012)



5.1.3 交叉制表（Cross-tabulations）

继续对照图 5-7. (Figure 5-7)的流程图，以下是问题 C007 和 C008A 的交叉列表，用来查看跳转的结果。此数据只供参考;以变量赋值标签显示。图 5-8(Figure 5-8)显示 48 位在 C007 回答“吸过”(YES) 的受试对象也回答了 C008A 关于从多大年龄开始吸烟的问题。但是在 C007 回答“从没吸过”(NO)或表示“不知道”(Don’t Know) 的 252 位，没有答 C008A，数据是“合理无答案”(N/A)。这个交叉列表显现了所有的受试对象都跳转正确。

图 5-8 (Figure 5-8). C007 和 C008A 的交叉列表

```
PROC freq data=temp;
Table c007*c008a /list missing;
Run;
```

The FREQ Procedure

C007	C008A	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	12	1	0.33	1	0.33
Yes	16	2	0.67	3	1.00
Yes	17	5	1.67	8	2.67
Yes	18	21	7.00	29	9.67
Yes	Unknown age	19	6.33	48	16.00
No	N/A	246	82.00	294	98.00
Don't know	N/A	6	2.00	300	100.00

图 5-9(Figure 5-9)是另一个交叉列表，用来查看 C008A 和 C008B 的跳转的结果。只有在 C008A 答不知道(Don’t Know) 多大年龄开始吸烟的问的受试对象才会被问到问题 C008B，关于开始吸烟的年份。换句话说，任何在 C008A 回答了年龄或者没答 C008A（就是在 C007 回答“从没吸过”(NO)或表示“不知道”(Don’t Know) 的 252 位，跳转过 C008A 的受试对象是不该回答 C008B 的。

图 5-9 (Figure 5-9). C008A 和 C008B 的交叉列表

```
Proc freq data=temp;
Table c008a*c008b /list missing;
Run;
```

The FREQ Procedure

C008A	C008B	Frequency	Percent	Cumulative Frequency	Cumulative Percent
N/A	N/A	252	84.00	252	84.00
12	N/A	1	0.33	253	84.33
16	N/A	2	0.67	255	85.00
17	N/A	5	1.67	260	86.67
18	N/A	21	7.00	281	93.67
Unknown age	2005	1	0.33	282	94.00
Unknown age	2007	1	0.33	283	94.33
Unknown age	2008	3	1.00	286	95.33
Unknown age	2009	8	2.67	294	98.00
Unknown age	2010	5	1.67	299	99.67
Unknown age	Unknown year	1	0.33	300	100.00

在这一系列的问题里，所有的问题都会汇集到 C012 （接触二手烟），意思就是所有受试对象都不会跳转过 C012。也不会像 C008A 和 C008B，依据在 C007 的回答影响跳转的结果。图 5-10(Figure 5-10)显示 C012 和 C007 的交叉列表中，有 3 个受试对象数据缺空。这表示数据有错，因为所有人都被问及 C012。

图 5-10 (Figure 5-10). C012 和 C007 的交叉列表

```
Proc freq data=temp;
Table c008a*c012 /list missing;
Run;
```

The FREQ Procedure

C007	C012	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	0	24	8.00	24	8.00
Yes	1	3	1.00	27	9.00
Yes	2	3	1.00	30	10.00
Yes	3	3	1.00	33	11.00
Yes	4	3	1.00	36	12.00
Yes	5	3	1.00	39	13.00
Yes	6	3	1.00	42	14.00
Yes	7	3	1.00	45	15.00
Yes	Missing	3	1.00	48	16.00
No	0	99	33.00	147	49.00
No	5	6	2.00	153	51.00
No	6	103	34.33	256	85.33
No	7	38	12.67	294	98.00
Don't know	5	6	2.00	300	100.00

当你收悉变量与变量间的关系以及变转模式时，可以一次交叉列表数个变量。如此能够阅览受试对象的回答如何分布于变量间。图 5-11(Figure 5-11)显示之前在 C012 的 3 个数据缺空的受试对象，在 C007 回答吸过香烟，在 C008A 回答从 12 或 16 岁开始吸烟，没答 C008B 因为跳转到 C012。或者，表示没吸过香烟的受试对象（C007=No）却表示几乎每天接触二手烟（5， 6， 7 天）。这比表示吸过香烟的受试对象（C007=Yes）接触二手烟 5,6,7 天的人还要多。如果这个结果令人惊讶，最好再用其他的交叉列表来深入探讨，来决定是否人工比对数据和问卷来验证数值。

图 5-11 (Figure 5-11). 数个变量的交叉列表 (C007, C008A, C008B,和 C012)

The FREQ Procedure

C007	C008A	C008B	C012	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	12	N/A	Missing	1	0.33	1	0.33
Yes	16	N/A	Missing	2	0.67	3	1.00
Yes	17	N/A	0	2	0.67	5	1.67
Yes	17	N/A	1	1	0.33	6	2.00
Yes	17	N/A	2	1	0.33	7	2.33
Yes	17	N/A	3	1	0.33	8	2.67
Yes	18	N/A	0	3	1.00	11	3.67
Yes	18	N/A	1	2	0.67	13	4.33
Yes	18	N/A	2	2	0.67	15	5.00
Yes	18	N/A	3	2	0.67	17	5.67
Yes	18	N/A	4	3	1.00	20	6.67
Yes	18	N/A	5	3	1.00	23	7.67
Yes	18	N/A	6	3	1.00	26	8.67
Yes	18	N/A	7	3	1.00	29	9.67
Yes	Unknown age	2005	0	1	0.33	30	10.00
Yes	Unknown age	2007	0	1	0.33	31	10.33
Yes	Unknown age	2008	0	3	1.00	34	11.33
Yes	Unknown age	2009	0	8	2.67	42	14.00
Yes	Unknown age	2010	0	5	1.67	47	15.67
Yes	Unknown age	Unknown year	0	1	0.33	48	16.00
No	N/A	N/A	0	99	33.00	147	49.00
No	N/A	N/A	5	6	2.00	153	51.00
No	N/A	N/A	6	103	34.33	256	85.33
No	N/A	N/A	7	38	12.67	294	98.00
Don't know	N/A	N/A	5	6	2.00	300	100.00

5.1.4 重要提示

几组个案数据输入后,可以立即开始数据的查阅/清理,查看频率分布和交叉制表。 要在牢固的电脑上进行这些步骤,以保证保密的调查信息不被其他工作人员使用。不可打印包含个人信息的数据,把实体文件锁在桌子或保险柜里。工作完成后,把文件切碎。

养成报告数据分析结果的习惯。即使没有发现问题,和其他人分享你的分析结果也很重要。举列重点报告初始个案的数据质量。传给项目领导阅览。

- 讲述应该采取的步骤来纠正过去的数据错误和避免以后的错误。

- 记录对调查指示,表格,数据输入,和数据编码库所做的修改。
- 描述你要使用的数据当案,记录详细的下载日期和个案数目。
- 你使用的 SAS 或其他分析软件所用的运作命令要提供给其他人。注明运作命令的状态。
- 如果你的进展因查看可能出现的问题而中断,描述你的发现。例如,“一些日记中报告的进餐次数尤其低,经过查看这些个案,我们发现几页日记的数据没有完全输入。这些日记的数据输入正在进行,将被包括在下一次分析当中。”

根据输入输入的速度, 如果有新的几组数据收集上来,你应当再次重复数据审阅。只要更改了问卷,给访谈对象的指示,编码系统,或者数据输入系统,就要进行数据审阅。例如,一份问卷的某个问题,第一稿时以前需要只准圈选一个答案,但是第二稿（数据收集 6 个月后因应的改变）可以允许圈选多个答案。查找后面这些问卷里对这个问题的多个答案。再者，数据输入人员可能没有输入一致性的代码（或是经过了重新培训，代码有所改变）。例如,将空白替换成“B” 或 “9”。 要确认以前的空白要用新的编码例如“B” 或 “9”来替代。这时该重复数据审阅，以保证规格前后一致。图 5-12(Figure 5-12)显示一个叫 SALEPRICE 的变量, “B”的代码出现了 3 次,“999”出现了 12 次, 空白(Frequency Missing)出现了 202 次。对于空白的处理竟然有三种方式,显然的必需经由数据清理来纠正这些错误。

图 5-12 (Figure 5-12). 用来评估缺省数据处理方法的调查问卷频率分布

```
proc freq data=temp_QC1;
tables SALEPRICE;
run;
```

The FREQ Procedure

{SALEPRICE} PAYMENT RECEIVED

SALEPRICE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
B	3	16.67	1	5.56
10	1	5.56	2	11.11
15	1	5.56	5	27.78
30	1	5.56	6	33.33
41	1	5.56	7	38.89
150	1	5.56	8	44.44
181	1	5.56	9	50.00
200	2	11.11	11	61.11
236	1	5.56	12	66.67
244	1	5.56	13	72.22
261	1	5.56	14	77.78
415	1	5.56	15	83.33
535	1	5.56	16	88.89
550	1	5.56	17	94.44
999	12	66.72	29	100.00

Frequency Missing = 202

练习 5-1. CFCS数据清理过程中可能有出现什么情况呢?

想想前面展示的 CFCS 调查工具和数据清理步骤。 数据管理人员在看数据输出结果的时候有什么需要注意的呢? (提示:

- 数据符合逻辑吗?
- 有重复数据吗?
- 数据格式一致吗?
- 一致性代码使用适当吗或者有使用一致性代码吗?
- 数据和数据编码库一致吗?
- 数据文件要进行转换吗?

- 使用流程图来查看问题跳跃的方式
- 使用散点图来查看相关性

数据文件的递交

做完几轮数据查阅之后,要确保先前发现的问题没有在后来的数据输入中出现。与此相关的是,要决定该用累积频率分布的方法查看迄今为止的所有数据,还是只查看上次审阅数据之后新输入的数据。累积频率分布花费的时间更长,但如果有问卷或者系统整体上的变化时,应当使用这种方法。 每个审阅和清理过的数据文件都要收集成一个整合的文件来递交。这些递交或许会在数据收集的末尾一次完成,也或许象下个部分描述的一样反复几次进行。

查验每次文件递交中需要包括的文件。除了各种数据文件,例如数据编码本,调查问卷复印件,数据输入指示,打印的前 100 个个案的频率分布,或者数据变量的内容。 每个个案都要有独特的个案编码或个案代码,而且包含需要递交的所有个案。可包括所有有完整数据的个案,只有部分数据的个案,或者数据库的一小部分。要包含所有必要的变量。这些可能包括预载数据(例如人口普查街区),过去纵向调查的问卷答案,访谈答案,编写提醒标识,计算过的变量(譬如, 根据对两个或更多的问题答案所得出的答案)。

对分析数据的用户来说,数据编码本是一个极有价值的资料。没有数据编码本,分析研究员会浪费很多时间,来弄明白调查问卷里的每个数字或符号指的是什么。他们可能会弄不懂数据,作出不准确的猜测,从而影响他们研究的质量。如果你能对每个递交的数据档案都提供相应的数据编码本,就会提高自己的效率,因为数据收集完成后对数据递交的问题就会少很多。

要决定如何递交档案,使用什么安全措施或加密方式。在美国,有些大型健康和教育的研究项目使用专门的 FTP 地址来保证负责机构和数据收集人员之间安全地共享数据。更常见的方式是使用 Winzip 之类的软件对数据文件进行压缩,使用密码保护的方式,并复制到 CD, DVD, 移动光驱,或 USB 驱动器上。 密码应该至少包含 8 位字码,既包括字母,数字和符号(例如&*@#$)。 每次你准备递交文件,都要重新修改密码。只将密码通过电话(而不是电子邮件)来告诉最少量需要知道密码的人。

递交数据前,给你机构的某个人或 NCI 或者 CDC 中的合作者进行一次模拟递交,以保证数据可以正常查看,档案可以正常打开。使用与你计划使用相同的递交方式 (例如刻到光盘上的经过加密并压缩的文字文件)。 确认你办公室的接受人能使用那个工具,阅读里面的内容,打开文件,将它们上传到电脑或共享的网络上。确认 Word 文档资料例如编码本,数据输入指示,问卷/表格,能够阅览和打印。经过一次成功的内部测试后,将文件递交,并确认文件顺利提交。 如有必要,根据接收人的指示修改递交方法或者文件的格式。

练习 5-2. 需要递交的CFCS数据文件包括哪些?

下面两个文件包含在食品日记数据的递交档案中。你怎样用数据字典来理解数据文件呢?(为什么 ZD2 竖列里有空格呢? 还有什么能帮你来理解这些数据?)

膳食日记的数据文件格式剪影:

ZDF	ZDG	ZDH	ZD1	ZD2	ZD3	ZD4	ZD5	ZD6	ZD7
20	7	34	1	饺子	1	2	面粉	11405	100
20	7	34	1				猪肉	81101	30
20	7	34	1				韭菜	44401	50
20	7	34	1				西葫芦	43218	150
20	7	34	2	甜瓜	1	1		66108	500
20	7	34	2	水	1	1		166202	100
20	7	34	3	面条	1	2	面粉	44105	200

数据字典(结构):

Variable n	Variable description (Variable label)		Variable type	Variable width	Variable value	Variable value label	
变量名称	变量描述 (变量标签)		变量类型	变量宽度	变量赋值	变量值标签	
	中文 Chinese	英文 English				中文 Chinese	英文 English
ZDF	核实开始时间 (分	Start time- minute	数字型	2	0-59		
ZDG	核实结束时间 (小	Stop time- hour	数字型	2	0-23		
ZDH	核实结束时间 (分	Stop time- minute	数字型	2	0-59		
ZD1	进餐时间	Meal Time	数字型	1	1	早餐	breakfast
					2	上午加餐	morning snack
					3	午餐	lunch
					4	下午加餐	afternoon snack
					5	晚餐	dinner
					6	晚上加餐	evening snack
ZD2	食物或饮料名称	Name of food and Beverages	字符型	8			
ZD3	食物来源	Food source	数字型	1	1	家里	home
					2	食堂	cafeteria
					3	餐馆	restaurant
					4	市售	market
ZD4	食物加工方式	Description and preparation	字符型	8	1	生食或未加工	raw
					2	蒸	steamed
					3	煮	boiled
					4	炒	stir-fried
					5	煎	fried
					6	烧烤	roasted
					7	炸	deep-fried
					8	红烧	braised
					9	炖	stewed
					10	卤	marinated
					11	烘焙 (干)	baked
ZD5	食物组成	Components of Food or Beverag	字符型	8			
ZD6	食物编码	Food Code	数字型	6		1-100 , 011201-227006	
ZD7	食物进食量	Amonut(g)	数字型	4			

6. 数据质量和管理的其他问题

练习 6-1. CFCS是如何来解决新出现的数据质量问题的？

CFCS 项目遇到这些数据的质量问题吗？ CFCS 是可以如何解决这些问题呢？

1. 采访者相异处（例如，每个数据采访人完成多少访谈？）
2. 调查问卷的设计（例如，在数据收集工具中选项“是/否“的顺序不一致）
3. 调查人员进行高技术工作（例如，转换熟的为生的成分；估计有几克）
4. 开放式回答的编码（如为“其他”回答数据建立一个编码结构的流程）
5. 将这些经常要做的事嵌入所有的步骤中
   - 安全存放,检索,和处理实体文档资料(如纸质文件)
   - 报告和记录