Departmental Overview

The Department of Statistics has historically led the development of analytical methods for major aspects of the RERF research program estimating radiation risk. These methods have also been applied by others to numerous radiation studies of other cohorts. This includes the development of tools and methods to flexibly estimate radiation-associated excess relative and additive risk of cancer and other adverse health outcomes and to handle various problem areas in the data collected on major RERF cohorts, such as missing data on various covariates, error in exposure ascertainment, and unknown failures to register incident cancer cases due to undocumented out-migration from cancer tumor registry catchment areas, among many other examples. We also develop or adapt statistical methods to analyze the longitudinal data generated by the Adult Health Study and more recently by the F1 Clinical Study.

We have evaluated or developed novel methods of sub-cohort sampling to maximize statistical power when studies cannot measure necessary covariates on the entire cohort, to address special issues related to variables that modify the radiation risk per unit dose or are intermediate steps on a causal pathway between radiation and a health outcome under study, and to accommodate correlated competing health outcomes that arise over long-term follow-up. We also develop many special methods to meet the requirements of RERF's basic-science research in genetics, immunology, radiation biology, and molecular epidemiology, notably applying new methods to analyze their high-dimensional data.

Unlike other departments, we engage in both consulting and research and devote a majority of staff effort to our collaborative role. Given the crucial importance of sound statistical advice in the design, analysis, and reporting of studies, the Department seeks to adopt a proactive approach to consulting, beginning with early involvement in study design, and estimation of the statistical power of potential studies. Our proactive approach is facilitated through research cluster activities associated with research project development. This is critical information in evaluating how effectively a given project will be able to address its proposed scientific question, which in turn impacts its value to the RERF mission. Our second major responsibility is to provide sound statistical analyses of data gathered by investigators in all the research departments.

Work in radiation dosimetry, including investigation into dose uncertainty, is by definition central to the RERF mission of evaluating the effects of ionizing radiation on human health. The Department implements RERF dosimetry systems and maintains a database of survivor dose estimates, presently using the DS02 system developed by a combined external and internal scientific working group, which requires developing important methods to interpolate and extrapolate dose estimates from the coarse and limited-distance output of the dosimetry system, to provide dose estimates for survivors with incomplete data on shielding, and to correct for uncertainty in the dose estimates, among other procedures. The Department also provides key statistical and dosimetric support to RERF projects in biodosimetry. Statistics staff continued to organize meetings of a binational working group of external scientists working on evaluating the improvements to dose estimates for specific organs and tissues that could be realized by new computational models of the human body and transport calculations versus the 35-year-old calculations currently used by DS02, and the Department will be actively involved in that project as the new calculations begin in the coming year.

FY2017 Achievements

In the past year, members of the Department of Statistics have performed independent and collaborative research on a variety of topics regarding statistical methodology and radiation dosimetry, which is reflected in papers published in peer-reviewed journals (20 for which at least one Department member was listed as an author, including 11 originating in our Department) and presentations at scientific meetings (13 originating in our Department). In addition, members of the Department have been active collaborators on RERF studies, at all stages from proposal development within the research clusters, including experimental design and sample selection, to data analysis and manuscript preparation. Below, we provide details on methodological and collaborative research projects within several select initiatives. Additional details are provided in respective lists of presentations and publications.

Long-term follow-up studies

Follow-up of the LSS, the AHS, the *in utero* cohort, and the F1 cohort is central to RERF's mission to quantify the effects of radiation exposure. We collaborate closely with researchers in Epidemiology and Clinical Studies, as well as the U.S. National Cancer Institute, on studies regarding mortality and morbidity among these cohorts. Over the past year, statisticians have collaborated on studies regarding the incidence of solid cancer at specific cancer sites, including breast (Brenner/French), lower digestive (Sugiyama/Misumi), liver (Sadakane/French), uterine (Utada/Cologne), and urothelial (JY Kim/Cologne) cancers, as well as pooled studies on leukemia (Little/French). These collaborations have motivated several methodological investigations regarding the statistical design and analysis of cohort studies, with focus on the shape of the radiation dose-response, including:

- Effect of follow-up period on minimal-significant dose (Cologne);
- Source of non-linearity in the radiation dose-response for solid cancer (Cologne);
- Selection of an appropriate reference group of unexposed individuals (French);
- Semi-parametric smoothing for joint effects of radiation and covariates (Misumi);
- Semi-parametric estimation of the radiation dose response at low doses (Furukawa);
- Impact of adjustment for pre-bombing population density (French);
- Efficient procedures for multiple imputation in large-scale studies (Furukawa).

We have developed and implemented novel methods for accommodating correlated competing health outcomes in aging RERF cohorts. Over the past year, we have implemented joint analyses that facilitate more detailed estimation of radiation risks and provide inference regarding differences in radiation effects, such as proximal versus distal colon cancer (Sugiyama/Misumi), liver versus intra-hepatic bile duct cancer (Sadakane/French), and different groupings of all-solid cancer (Cologne). In addition, we have implemented joint regression models for longitudinal and survival outcomes in AHS studies regarding red blood cell distribution width (Yoshida/Misumi), lymphocyte and monocyte counts (Yoshida/French), and blood pressure (I. Takahashi/Furukawa). We have continued research on joint regression models for longitudinal data with outcome-dependent observation times (French), and have designed a competing risk survival analysis for an AHS study of incident cardiovascular diseases (I. Takahashi/French).

We have also developed approaches for designing studies within these long-term follow-up studies, which are intended to reduce logistical effort and consumption of precious biological specimens, without suffering substantial loss of statistical efficiency, such as counter-matched nested case-control studies and stratified case-cohort studies. We have begun to consider the use of exposure-enriched case-control studies for cross-sectional sampling in studies aimed at causal modeling or integrated analyses of molecular endpoints. Methods of analyzing data collected under these designs are constantly being updated, such as with two-phase analysis of nested case-control and case-cohort designs. We are keeping abreast of the evolution of these methods and one of us (Cologne) has been collaborating with outside researchers in this area.

Causal inference

In an observational study such as the LSS, treatment assignment is typically assumed to be effectively random conditional on measured covariates. However, the presence of unmeasured confounding can result in non-random treatment assignment, such that standard analysis methods can provide biased estimates of treatment effects. The potential for measured and unmeasured confounding motivates consideration of sensitivity analyses to assess how much bias, due to non-random treatment assignment, would be necessary to change the conclusions. We initiated a collaboration to develop new methods for performing sensitivity analyses based on evidence factors (French, with D Small). We developed approaches for combining the evidence from multiple evidence factors in a single observational study, while controlling the family-wise error rate. We showed that combining the evidence from multiple evidence factors is more efficient than considering each factor separately.

Questions regarding the mechanisms of radiation effects can be addressed by quantifying the extent to which radiation-outcome associations are mediated by covariates that capture clinical or biological information. We evaluated sample-size requirements and statistical power for generating inference regarding the mediation proportion in studies of mediation (Cologne/Cullings, with YM Kim); this was one of the most comprehensive assessments to date of power for assessing mediation with computer-intensive methods such as the bootstrap. We initiated work on estimating the mediation proportion with data collected under the counter-matched, nested case-control study design (Cologne/Cullings, with YM Kim); this was the first examination of methods of inference for mediation in the nested case-control study design, and we showed that selecting controls by counter matching does not impact inference on mediation. We completed an analysis of atherosclerosis risk involving latent clinical symptoms of atherosclerosis estimated with clinical indicators (I. Takahashi/Cologne). We found that radiation is associated with latent factors for arterial calcification and plaque, even though it was not significantly associated with their clinical indicators individually, demonstrating that causal models and structural regression with latent factors can have greater power than ordinary regression analyses applied to a large number of individual endpoints.

Mechanistic models

We recognize the important need to complement clinical and epidemiological studies regarding radiation health effects with biologically-based models to elucidate the mechanisms of radiation-related carcinogenesis. A collaboration with the Institute of Radiation Protection, Helmholtz Zentrum Munchen (HGMU), Germany, was continued on biologically-based mechanistic modeling of radiation-related carcinogenesis for lung cancer and thyroid cancer. A manuscript on mechanistic modeling of the roles of smoking and radiation in molecular

pathways to lung adenocarcinoma has been developed (Furukawa, with JC Kaiser). We identified a new model that accurately reproduced the incidence of lung adenocarcinoma stratified by risk factor and molecular pathway, confirmed the existence of the two pathways, explained population differences in driver mutation frequencies, and provided the first direct epidemiological linkage of risk factors (smoking and radiation) with the two molecular classes of lung adenocarcinoma. An international workshop on biologically-based modeling was co-organized with HGMU, to be held at RERF in March 2018.

High-dimensional data

Future RERF research requires maximizing the potential of longitudinal biosamples collected among the clinical cohorts, along with the application of state-of-the-art technologies in genomics, proteomics, and metabolomics. We collaborated with the Biostatistics and Informatics Resource at the University of Hawaii Cancer Center to implement integrated analysis of expression (RNA-seq) and methylation data to jointly analyze aging effects on these outcomes in naïve CD4-positive T lymphocytes (Cologne/Misumi, with K Yoshida). Although the result was negative, it adds to our growing information that naïve CD-4 positive T cells might not demonstrate strong effects of aging. This project included the writing of R programs to do preliminary summarization and processing of the large data matrices, which will form a basis for future integrated-omic studies at RERF.

Radiation dosimetry and dose error

Radiation dosimetry, including investigations into dose uncertainty, is by definition central to the RERF mission of evaluating the effects of ionizing radiation on human health, as such efforts provide individual dose information for measures of dose-response. Over the past year, we investigated the (limited) information on the relative biological effectiveness of neutrons in the LSS solid cancer incidence data using the latest follow-up with DS02 and DS02R1 dose estimates, which had not been attempted since the introduction of DS86 (Cordova/Cullings). We continued work on spatial analyses of thermoluminescent dosimetry measurements of bricks and tiles, which are legacy data compiled for the DS02 report, regarding questions raised in the literature about potential use of those data to address controversial assertions by some investigators of undocumented residual radioactivity from the Hiroshima atomic bomb (Cullings/Grant). We began analysis of a new set of fluorescent in situ hybridization (FISH) data on chromosomal aberrations, in relation to DS02R1 dose estimates, to be analyzed jointly with a revised and partially overlapping set of conventional Giemsa-staining data that were previously analyzed in a landmark 2001 paper using DS86 dose estimates (Y. Kodama/Nakamura/Cullings).

We had two major publications on dose uncertainty. First, we quantified the potential for bias in radiation risk estimates obtained from regression calibration correction for dose error, and proposed an alternative method (SIMEX) that avoided an assumption regarding the distribution of unobservable true doses, in contrast to the regression calibration method currently in use (Misumi). Second, we provided an assessment of dosimetric uncertainty without parametric assumptions by using chromosomal aberration data available for a subset of the LSS as an instrumental variable, concluding that the attenuation of the radiation risk estimate in the LSS by dosimetric uncertainty (classical error) might be greater than previously assumed (Wang/Cullings). We continue to work both autonomously (Cullings,

Misumi) and collaboratively with outside investigators (Cullings, with D. Pierce) in related areas.

Training and mentoring

In addition to research, we have important roles in training, both as instructors and learners. In the past year, we provided lectures on radiation risk modeling for the RERF international training program, a two-day series of lectures on environmental statistics at Kurume University, and a one-day short course on joint regression models for longitudinal and survival data at Korea University. We are active participants in the Hiroshima Statistics Study Group, which is held monthly at RERF in collaboration with Hiroshima University. Regular department seminars allow researchers to obtain feedback on work in progress and facilitate intra-departmental collaboration. We will host an international workshop on biologically-based modeling in March 2018.

RERF has instituted rather rigorous periodic reviews for fixed-term research scientists, which includes junior members of the Department. At the Department level, we created an electronic tracking system for consulting and collaborative work that will help to insure equitable allocation of related workloads within and between members of the Department.