



Cancer Research Center Hotline

The Importance of Biostatistics to Cancer Research

Lynne R. Wilkens DrPH
Associate Specialist – Biostatistician
and
Leo Wang-Kit Cheung PhD
Assistant Specialist – Biostatistician
Cancer Research Center of Hawaii

Overview. Statistics is defined by the Merriam-Webster Collegiate dictionary as “a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.” In practice, statistics is multi-disciplinary, in that it uses the tools of mathematics and computing to answer questions in other fields. Statisticians are members of research teams, who must become knowledgeable about the field of application, in order to understand the data issues. Statisticians are committed to the scientific method, believing that well-designed research studies with good data collection techniques are essential for uncovering the truth. Statisticians are specialists in understanding the nature of uncertainty in data and accounting for it in the design and analysis of research studies. Biostatistics involves the development and application of statistical techniques specifically to scientific research in biological sciences and health-related fields, including medicine, epidemiology and public health.

Biostatisticians are involved in every step of a research project. They help formulate scientific questions to be answered, determine appropriate sampling techniques, coordinate data collection and management procedures, and carry out statistical analyses to answer the scientific questions. They ensure that the data meet the assumptions of the statistical model or test being used and develop new statistical techniques as needed. Biostatisticians use the tools of data base management and statistical programming.

Specific areas of study in cancer research that require statistics include clinical trials and group intervention trials, epidemiology, environmental science, genetics and molecular biology.

Clinical Trials and Group Intervention Trials. Clinical trials in cancer research are experimental studies where patients are, in general, randomly assigned to receive different cancer treatments. Commonly, a new treatment is compared against the standard care for the particular type of cancer under study. Researchers may hypothesize that a new therapy improves survival or is safer for the patient. Pharmaceutical companies use clinical trials to determine the effectiveness of new therapies, and the National Cancer Institute uses the results of national clinical trials to make recommendations on standard care for cancer patients. CRCH’s Clinical Trials Unit coordinates these national clinical trials for the patients in Hawai‘i. Statisticians are involved in the design and analysis of these experimental studies. Determining the appropriate sample size for clinical trials is very important, as a study that is too small will not be able to answer the question of interest. Also, a study that is larger than

needed may expose more patients than necessary to an experimental therapy, before safety issues are adequately understood. Statisticians develop randomization schemes to help ensure that the treatment groups are similar in important prognosis factors, other than the therapy under study. If treatment groups differ on important prognosis factors, such as age, the question of which therapy is better will be more difficult to answer. Randomization to treatment generally results in groups that are similar in age. Adaptive randomization schemes have been developed for trials with multiple treatment doses to minimize the number of patients randomized to more toxic doses.

Statisticians plan and execute the analysis of clinical trials. The analysis generally involves determining the relationship between the outcome variable and the assigned treatment group. However, certain factors complicate the analysis. Clinical studies often recruit patients from multiple clinics and require multiple measures from patients over time. The analysis of the data must then account for the correlation structure between the data points. Also, patients may withdraw from longer studies. If the data collection mechanism that leads to missing data is related to the treatment or outcome, the analysis must account appropriately for the missing data to derive the correct result. For example, sicker patients may drop out of a study, resulting in the final outcome being available only for patients with a good prognosis. Comparing the treatments only for patients with final outcomes may not give an accurate picture of the benefits of the therapies. Biostatisticians have studied the effect of such missing data and have developed special analysis techniques for dealing with missing data.

Group intervention trials are studies that randomize communities or groups of individuals to an intervention under study. CRCH’s Social and Behavioral Science Program uses these studies to investigate ways to effect behavior change in smoking, diet and sun protection practices at the population level. Analysis of group intervention trials is complicated by the challenge of measuring behavioral change in the community and the fact that communities are randomized to interventions, but measurements are made at the individual level.

Epidemiology. Epidemiology is the study of the distribution and determinants of human disease. CRCH’s Epidemiology Program studies lifestyle and genetic risk factors for cancer in Hawai‘i’s multiethnic population.

In descriptive epidemiology, the extent of the disease problem is measured. The Hawai‘i Tumor Registry of the CRCH reports cancer incidence and mortality rates for the state by sex and ethnicity. Biostatisticians ensure the accuracy of these rates, by comparing the ways data, such as residency and ethnicity, are collected for the numerator and denominator, and making adjustments to ensure that the definitions are similar. Analytic epidemiology uses observational studies to answer scientific questions. Standard research designs are cross-sectional studies, case-control studies and cohort studies. All require appropriate choice of participants, so that the results can be generalized to the larger population that the participants are meant to represent. In observational studies, the relationship between exposures of interest and disease is studied. However, because individuals are not randomly assigned to the exposure, the analysis of these studies is complicated by potential confounding

from other risk factors. Statistical techniques are used to control for the effect of other factors. For example, a researcher may study the possible protective role of vegetable consumption in the development of lung cancer. Cigarette smoking is known to be a strong risk factor for lung neoplasm, and it has been found that heavy smokers are less likely to eat vegetables. Therefore, careful controlling for tobacco exposure is needed to determine if vegetable consumption truly protects individuals from lung cancer, or is simply a surrogate measure for the extent of tobacco exposure. Another statistical challenge in observational, and even clinical studies, is that most exposures can not be measured precisely. Most statistical models, however, assume that independent variables are measured without error. In the above example, the long-term consumption of vegetables and exposure to tobacco is usually self-reported by the study participants; this self-report is prone to error in recall. Statisticians determine the level of error in the measurement of exposures and its effect on the results. Special study designs and models have been developed to deal with measurement error in exposure variables.

Environmental Science. Environmental science uses statistics to characterize potential adverse health effects resulting from exposure to environmental hazards. Both the level of exposure present, and the relation between exposure and incidence of disease, are quantified. As an example, the Environmental Protection Agency monitors the level of pollutants in the air and water, and estimates the level of health risk posed. Measuring the level of exposure to environmental hazards in the general population is difficult. For instance, air pollution levels may vary by geographic features, such as proximity to mountains, and by time of day. Biostatisticians determine appropriate sampling methods to determine how best to measure overall exposure.

Genetics and Molecular Biology. Genetics and Molecular Biology are becoming an important component of cancer research. With advances in biotechnology, we can now study structures and functions of genes and particular DNA regions, and their organizations in the human genome. It is now known that certain regions of DNA in our genome play various regulatory roles in controlling human body functions by promoting or silencing the expression of genes. The understanding of DNA and the networks of genes in turn gives us important knowledge of the molecular nature of the development of many diseases such as cancer. Discovering exact functions of various regions of DNA and confirming the associations and/or causations between genetic variations and diseases are the challenges of the Human Genome Project. The completion of the human genome — the genetic blueprint of humankind — has provided researchers large amounts of DNA data. With about 3.1 billion units of DNA in the human genome being available as sequence data, the need to process the information, to reveal hidden patterns and to interpret encoded messages has created new problems that are mostly statistical in nature. A newly emerged science termed bioinformatics is dedicated to the development and application of mathematical, statistical and computational methods in genome research. It certainly falls within the realm of biostatistics, and has also been identified as a part of the field of statistical genetics in the Handbook of Statistical Genetics.

Genetics has already started to be incorporated in cancer research

studies. In clinical studies, treatments are being targeted to patients based on gene expression profiles of their tumors. In epidemiology, the role in cancer risk of interactions between multiple genes and between genes and the environment are being investigated. There are a few genotypes that have been found to have a large effect on the risk of cancer, such as a mutation in the BRCA1 gene that has been linked to the development of breast and ovarian cancer. However, these genotypes are generally rare. Susceptibility genes, which affect an individual's ability to activate or detoxify a carcinogen, are being investigated for their role in cancer development. Most genotypes under study are common and would presumably only affect cancer risk in the presence of a carcinogen.

Future challenges. Cancer research studies are becoming ever more complex, collecting data from multiple sources with varying accuracy. Aggregating and interpreting this data to provide an accurate picture of the underlying phenomenon will require well-trained specialists in biostatistics. In addition, these individuals will be needed to support the projected growth of biomedical research in Hawai'i. However, there is a nationwide shortage of biostatisticians as noted in The Seventh Report to the President and Congress on the Status of Health Personnel in the United States and Objectives for the Nation. The lack of any degree-granting biostatistics program in Hawai'i makes it especially difficult to find people with these skills. The reinstatement of the graduate program in biostatistics at the University of Hawai'i, as well as development of additional statistical training at the university level, would help meet this demand as well as provide a good career option for people in Hawai'i.

For more information, please visit the Cancer Research Center's website at www.crch.org.



Aloha Laboratories, Inc.
...When results count

**CAP accredited laboratory
specializing in Anatomic
Pathology
Quality and Service**

**David M. Amberger, M.D.
Laboratory Directory**

Phone: (808) 842-6600

Fax: (808) 848-0663

**E-Mail: results@alohalabs.com
<http://www.alohalabs.com>**