<table>
<thead>
<tr>
<th>Monday, July 27</th>
<th>Tuesday, July 28</th>
<th>Wednesday, July 29</th>
<th>Thursday, July 30</th>
<th>Friday, July 31</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Monitoring Committees and Statistical Data Analysis Centers: Guiding Principles and Best Practices (Day 1 of 2)</td>
<td>Data Monitoring Committees and Statistical Data Analysis Centers: Guiding Principles and Best Practices (Day 2 of 2, morning only)</td>
<td>Bayesian Biostatistics</td>
<td>Use of Prognostic &amp; Predictive Biomarkers in the Design of Clinical Studies</td>
<td>Bayesian Adaptive Clinical Trial Design</td>
</tr>
<tr>
<td>Introduction to Survival Analysis</td>
<td>Survival Analysis for Observational Studies</td>
<td>Survival Analysis for Clinical Trials</td>
<td>Introduction to Longitudinal Data Analysis</td>
<td>Joint Modeling of Longitudinal and Survival Data</td>
</tr>
<tr>
<td>Modern Statistical Learning for Observational Data (Day 1 of 2)</td>
<td>Modern Statistical Learning for Observational Data (Day 2 of 2)</td>
<td>Small Area Estimation</td>
<td>Age-Period-Cohort Modeling and Analysis</td>
<td></td>
</tr>
</tbody>
</table>

Scroll down for module descriptions.
Introduction to Survival Analysis – 8:30 a.m. – 5 p.m.
Instructors: Brown, Elizabeth; Chen, Ying

Censored time-to-event data, where not all subjects experience the event of interest, are common in clinical and epidemiologic research. Examples include randomized controlled trials of therapies for cancer and other chronic diseases, comparative effectiveness research, and epidemiologic cohort studies. This module provides an introduction to censored time-to-event data and classical survival data analysis methods used in biomedical studies.

In this module, we will provide examples of studies where survival analysis is used and where it should not be used. We will describe how incomplete data on time-to-event outcomes (censoring) occurs. We will introduce important functions, including the survival function, the hazard function, and the median survival time, and show how censored data can be used to estimate them and compare the time-to-event experience between groups.

The module will explain key concepts unique to survival analysis such as risk sets and informative censoring. It will introduce the Cox regression model, and show how to examine the proportional hazards assumption.

The course will focus on application and understanding the concepts with examples from the biomedical literature; mathematical details will be kept to a minimum.

Modern Statistical Learning for Observational Data (Day 1 of 2) – 8:30 a.m. – 5 p.m.
Instructors: Benkeser, David; Carone, Marco Carone

While clinical trials provide the highest level of evidence to compare clinical treatments or public health interventions, they are often not feasible due to ethical, logistic or economic constraints. Observational studies provide an opportunity to learn about the effect of interventions for which little or no trial data are available. These studies constitute a potentially rich and relatively cheap source of information. However, in such studies, treatment or intervention allocation may be strongly confounded by other important patient characteristics and much care is needed to disentangle observed relationships and infer causal effects.

In this course, we will provide an overview of modern statistical techniques for analyzing observational data. We will focus primarily on recent advances in the field of targeted learning, which facilitates the use of state-of-the-art machine learning tools to flexibly adjust for confounding while yielding valid
statistical inference. In contrast, conventional techniques for confounding adjustment rely on restrictive statistical models and may, therefore, lead to severely biased inference. Use of the Super Learner framework, an implementation of model stacking, will be discussed as a particularly appealing means of performing flexible, pre-specified adjustment for confounding.

We will discuss methods for comparative effectiveness studies for single time-point interventions. We will also introduce the multi time-point extension of these methods and discuss strategies for dealing with missing data. Methods will be illustrated using data from recent observational studies and extracted from electronic medical records. Analyses will be illustrated in R but knowledge of R is not required for this module. In addition to lectures, the course will include in-class, hands-on activities to allow students to familiarize themselves with the methods and tools.

The two-day course is geared towards health science researchers with at least basic experience in data analysis and statistics. A basic understanding of the following concepts would be helpful: confounding, probability (e.g., what is meant by the distribution of random variable, its mean and its variance), statistical inference (confidence intervals, hypothesis tests), and regression (linear and logistic). Advanced knowledge of these topics is useful, but not necessary. Equivalent UW SPH course pre-requisites are BIOS 511/512 (or BIOS 514/515).
Tuesday, July 28

**Data Monitoring Committees and Statistical Data Analysis Centers: Guiding Principles and Best Practices (Day 2 of 2) – 8:30 a.m. – Noon**
Module description coming soon.
Instructors: Ellenberg, Susan; Fleming, Thomas; Wittes, Janet

**Topics in Clinical Trials: Surrogate Endpoints & Random High Bias Exploratory Analyses (Half-day stand alone module) – 1:30 - 5 p.m.**
Module description coming soon.
Instructor: Fleming, Thomas

**Modern Statistical Learning for Observational Data (Day 2 of 2) – 8:30 a.m. – 5 p.m.**
Instructors: Benkeser, David; Carone, Marco Carone
See Monday, July 27, for module description.

**Survival Analysis for Observational Studies – 8:30 a.m. – 5 p.m.**
Instructors: Brown, Elizabeth; Chen, Ying

This module will cover advanced topics in survival analysis, with an emphasis on applications in studies relying on observational data, including epidemiologic cohort studies and comparative effectiveness studies. This module will:

- Describe Cox regression models suitable for examining adjusted associations and effect modification;
- Cover the important choice of the time scale for the analysis, and discuss how to analyze data on subjects who enter observation after time zero (left entry and left truncation);
- Cover methods for appropriate inferences when there are competing risks, including the Cox regression model for cause-specific hazard functions and the Fine-Gray model for hazards related to the cumulative incidence function;
- Discuss biases that can arise in observational data: confounding, immortal time bias and index event bias, and how to treat them in the analysis, and
- Cover issues related to time-dependence in the Cox regression model, including how to incorporate hazard ratios, exposures, and adjustment variables that depend on time using time-dependent coefficients, time dependent covariates, and time-dependent stratification.

The course will focus on application and understanding the concepts with examples from the biomedical literature; mathematical details will be kept to a minimum. Knowledge of material in Introduction to Survival Analysis module will be assumed.
Evaluation of Biomarkers and Risk Models – 8:30 a.m. – 5 p.m.
Instructor: Kerr, Kathleen

This module covers methodology for evaluating biomarkers and risk prediction models, covering principles, concepts, metrics, and graphical tools.

We will discuss motivations for risk prediction in clinical medicine and public health, and clarify the concept of “personal” risk. Metrics and graphical tools will include ROC curves and AUC; calibration plots for risk prediction models; and net benefit and decision curves. The module will also discuss methods for comparing risk prediction models and, in particular, assessing the prediction increment of a new biomarker. We will also consider evaluating the utility of a single or composite biomarker for prognostic enrichment of a clinical trial.

There will be an opportunity for hands-on practice in R using relevant packages such as rms, rmda, and BioPET. However, the software component of the module is small and knowledge of R is not required for this module.
Bayesian Biostatistics – 8:30 a.m. – 5 p.m.
Updated title and description coming soon.
Instructor: Inoue, Lurdes

Survival Analysis for Clinical Trials – 8:30 a.m. – 5 p.m.
Instructors: Brown, Elizabeth; Chen, Ying

This module will cover advanced topics in survival analysis, with an emphasis on applications in randomized clinical trials (RCTs) with censored time-to-event outcomes. The module will:

- Review the logrank test and introduce testing procedures that weight group comparisons differently over the follow-up time interval;
- Introduce the restricted mean survival time and tests to compare it between groups;
- Review methods suitable for examining the adjusted association between an explanatory variable and a censored event-data outcome;
- Cover how information is accrued when there is group-sequential monitoring, and
- Cover power and sample size computations for an RCT with censored time-to-event outcomes.

The course will focus on application and understanding the concepts with examples from the literature; mathematical details will be kept to a minimum. Knowledge of material in the Introduction to Survival Analysis module will be assumed.

Methods for Developing and Evaluating Prediction Models for Dynamic Decision-Making – 8:30 a.m. – 5 p.m.
Instructors: Bansal, Aasthaa; Heagerty, Patrick

Many medical decisions involve using accumulated information on patients under surveillance to predict transitions in future health status, such as progression of disease or advancement to death. Longitudinal studies allow investigators to correlate changes in time-dependent exposures or biomarkers with subsequent health outcomes. At any given time, an individual’s longitudinal measures up to that time may be used to update the predicted risk of future adverse outcomes and guide medical decisions regarding monitoring and treatment. For example, high-risk individuals may be targeted for preventive strategies or aggressive treatments, whereas less frequent follow-up may be recommended for low-risk individuals.

In this course, participants will learn:

(1) Flexible approaches, such as joint models and partly conditional models, for modeling dynamic prediction rules for risk of a future adverse outcome using longitudinal trajectories up to the time of prediction, and
(2) Methods for evaluating predictive performance using summary measures that are appropriate for censored survival outcomes, with a focus on predictive accuracy using time-dependent sensitivity and specificity for prognosis of a subsequent event time.

Methods will be illustrated using examples from HIV, end stage renal disease, and organ transplantation settings. The course will include hands-on training and demonstration of relevant R packages for answering research questions. Real-data examples for analysis will be provided and the instructors will discuss implementation and interpretation.

This course is designed for those with basic understanding of methods for correlated data and survival analysis. Some experience with programming in R is preferred, but not required.
Many human diseases are heterogeneous in pathogenesis, prognosis and response to treatment. Nevertheless, clinical trials generally provide eligibility for a wide range of patients who often cannot reasonably be expected to respond similarly to molecularly targeted treatments. Lack of benefit for the majority of eligible patients can mask benefit for subsets of patients unless there are mechanistic biomarkers for subdividing the heterogeneous population.

Oncology has made substantial therapeutic progress using biomarkers which indicate particular somatic mutations which drive disease invasion. This has dramatically changed the process of drug development and evaluation with the development of drugs specifically to inhibit the activated protein product of such mutations. In these cases the enrichment design in which only biomarker positive patients are enrolled has been used in the past decade for the rapid regulatory approval of large numbers of new oncology drugs. In many cases, however, the biology of the disease is more complex and an appropriate predictive biomarker is not known in advance.

In this session, we will review clinical trial designs for the development of new therapeutics and companion diagnostics to inform their use. These include settings in which there is a candidate biomarker but the biological evidence for excluding biomarker negative patients is not compelling. In some cases, the relevance of the biomarker is established but an appropriate cut-point for positivity is only imprecisely known based on previous studies. In other cases, there are multiple candidate biomarkers and an effective classifier utilizing all or some of them needs to be developed. Using carefully constructed adaptive designs and re-sampling methods, we can sometimes both adaptively develop an effective classifier and internally validate it for identifying a subset in which the drug is effective in a single phase III trial. We will describe adaptive enrichment designs in which eligibility is modified adaptively during the trial using pre-specified frequentist or Bayesian strategies which preserve the type I error level of the trial while dramatically increasing the statistical power. Other designs to be discussed include the adaptive signature design, adaptive threshold design, basket designs, platform designs and a new class of run-in designs which use pharmacodynamics biomarkers as predictive markers. We will also discuss the development and use of prognostic signatures to facilitate the selection of an appropriate control group to avoid a high placebo response rate. Finally, we will discuss the use of observational and “real world” data to refine the intended use population for previously approved drugs.

The class is targeted to statisticians and others interested in clinical trials who want to become aware of the new ways of dealing with heterogeneous diseases to find effective treatments. It is for individuals interested in new designs and paradigms for development of treatments and for finding the most appropriate intended use population. The designs discussed are adaptive in the definition of the target patient population. Extensive use will be made of re-sampling methods for de-biasing re-substitution estimates and on pre-validation. We assume that the student is familiar with the standard features of randomized clinical trials as emphasis will be on new aspects.
Introduction to Longitudinal Data Analysis – 8:30 a.m. – 5 p.m.
Instructor: French, Ben

Longitudinal studies follow individuals over time and repeatedly measure health status, which facilitates prospective ascertainment of exposures and incident outcomes, and identification of changes over time within individuals. Analyses of longitudinal data must account for the correlation that arises from collecting repeated measures on the same individuals over time.

This module will overview statistical methods for the analysis of longitudinal data, with a focus on regression-based methods such as generalized linear mixed-effects models and generalized estimating equations. Relevant theoretical background will be provided. An illustrative example (conducted in R) will be used to illustrate analysis approaches, modeling strategies, and interpretation of results. This course is targeted toward individuals with little or no prior experience with statistical methods for longitudinal data analysis; experience with linear and logistic regression would be useful.

Propensity Scores – 8:30 a.m. – 5 p.m.
Instructor: Stephens, David

The propensity score is a key component of many causal inference procedures. After establishing the basic causal inference framework, we will outline the key methods of construction of propensity score functions, and study their core mathematical properties. We will detail the use of the propensity score in matching, inverse weighting and regression adjustments that allow the unconfounded effect of an exposure or treatment of interest to be estimated consistently.

Using the framework of semiparametric inference, we will contrast the statistical properties of estimators derived using each method. We will investigate issues of model selection for the propensity score, and demonstrate the utility of judicious choice of predictors that enter into the propensity function. This will be illustrated in standard problems and also in the case of high-dimensional predictors. Longitudinal data will also be studied in the causal setting.

Finally, we will develop the Bayesian framework for handling causal inference and investigate how propensity function construction and usage translates to the new setting.

All methods will be illustrated using examples from biostatistics, health research and econometrics. Computation will be performed in R.

Small Area Estimation – 8:30 a.m. – 5 p.m.
Updated title and module description coming soon.
Instructor: Wakefield, Jonathan
Bayesian Adaptive Clinical Trial Design – 8:30 a.m. – 5 p.m.
Instructor: Connor, Jason

The course will introduce Bayesian adaptive designs and cover numerous examples ranging from early to late stage trials. It will describe the range of potential adaptations and include operational benefits and challenges.

The course will introduce students to the skills and considerations necessary to construct such designs, including how to select design parameters based upon simulating trial operating characteristics. We will learn to compare designs to one another and fixed on numerous performance metrics.

The course will illustrate Platform trials involving multiple drugs, Goldilocks trials for pivotal trials using predictive probabilities for sample size selection, and dose-finding studies. We will also discuss operational aspects and challenges that arise when conducting an adaptive trial.

Some knowledge of clinical trials and statistics is necessary, though a Bayesian background is not.

Joint Modeling of Longitudinal and Survival Data – 8:30 a.m. – 5 p.m.
Instructor: French, Ben

Longitudinal studies follow individuals over time and repeatedly measure health status. Analyses of longitudinal data are often complicated by several factors that can threaten the validity of standard analysis methods. First, missing data in longitudinal outcomes can arise when individuals are lost to follow-up, either due to drop-out (e.g., in randomized trails) or death (e.g., in long-term observational studies). Second, when modeling intermittently measured time-dependent covariates in a survival analysis, biological variation can lead to measurement error. Joint modeling of longitudinal and survival outcomes has emerged as a novel approach to handle these issues.

We will detail the use of mixed-effects models for the analysis of repeated longitudinal measures, Cox regression models for the analysis of event-time outcomes with longitudinal measures as time-dependent covariates, and their combination in a joint modeling framework. An in-depth data analysis (conducted in R) will be used to discuss analysis strategies, the application of appropriate analysis methods, and the interpretation of results.

Developing Prognostic and Predictive Biomarkers with High Dimensional Data – 8:30 a.m. – 5 p.m.
Instructors: Simon, Noah; Simon, Richard

Modern medicine has graduated from broad spectrum treatments to molecularly targeted therapeutics. A major medical challenge is to develop companion diagnostics that provide guidance in directing use of these drugs to patients whose prognosis warrants them and whose disease is likely to be responsive to them. With continued advances in high throughput biotechnology, it is feasible to run whole genome assays on disease tissue whose features can be developed into prognostic and predictive biomarkers.
Specialized statistical methods are required for the development and evaluation of multivariate biomarkers using such assays because the number of candidate features is generally much greater than the number of patients. The goal of this module is to introduce ideas in high dimensional predictive modeling, to discuss model validation and testing, and to give hands-on experience with these tools to build predictive biomarkers on real, high-dimensional datasets.

Through the course, participants will become familiar with various methods in penalized regression, applications of cross-validation, as well as ideas in multiplicity and selection bias. Participants will gain experience applying these methods in R.

**Age-Period-Cohort Modeling and Analysis – 8:30 a.m. – 5 p.m.**
Instructor: Wakefield, Jonathan

In general, examination of disease rates may be carried out on three time scales: age (of the individual), period (time of diagnosis) and cohort (time of birth). Given any two of age, period, cohort, however, determines the third, and so one cannot uniquely identify the three different components.

Despite numerous warnings (Clayton and Schifflers, 1987a,b, Carstensen, 2007, Smith and Wakefield, 2017) over interpretation continues to occur in the literature.

In this course, the identifiability will be examined, and approaches to inference will be described. In particular, what can and what cannot be deduced from the data alone will be discussed.

Both frequentist and Bayesian methods will be presented. Throughout, ideas and modeling will be illustrated with examples. The examples will use publicly-available data, with methods implemented in the R programming language, with code provided, so that participants in the course will be able to carry out analyses with their own data.


