


[Home](#)
[Contact](#)
[UW Home](#)
[UW School of Medicine  
and Public Health](#)
[Site Map](#)

## Biostatistics and Medical Informatics Methodological Research Overview

[printable pdf page](#)



*UW Madison, Research Lab*

This page provides a brief overview of the types of research performed by our Department. Please consult each individual [Department Program](#) for more detailed information on each research area (if available).

[Research](#)
[Mission](#)
[Collaborative](#)
[Methodological](#)
[Research Grants](#)
[Software](#)
[Tech Reports](#)
[Research Resources](#)

**Biostatistics:** please see the [Biostatistics Program](#) for more specific information on areas of Biostatistical research. Examples include analysis of longitudinal/functional data, bayesian methods, clinical trials methodology, semiparametric inference and smoothing, sequential design of clinical trials, statistical computing, statistical genetics, and survival analysis methodology.

**Clinical Trials:** please consult the [Clinical Trials Program](#) page for more information on methodological research in the area of clinical trials, including a publication listing of representative works for each research area.

### Cancer Informatics Shared Resource

The faculty and staff of the Cancer Informatics Shared Resource (CISR) engage in collaborative projects with UWCCC researchers and improve and extend the computational resources, services, and capabilities available to support the UWCCC research enterprise.

## Collaborative Research

Medical Informatics Faculty engage in collaborative research in the following areas:

**Genetics/Genomics:** (Craven, Kendzierski, Newton, Page, Shavlik). Projects in this area include characterization of the genetic factors involved in susceptibility and resistance to cancer growth, mapping gene locations, analyzing genomic sequence data, analysis of mutagenesis experiments, analyzing gene-expression array data, optical mapping of genomes, predicting genomic regulatory elements, algorithms for analysis of mass spectrometry data, and novel gene-expression measurement technology. Collaborators include the UWCCC, McArdle, Computer Sciences, Pharmacy, Genetics, Biochemistry

**Clinical and Bioinformatics:** (Brennan, Craven, Page, Shavlik, DeMets). Collaborative projects include computational diagnosis of breast cancer including the determination of which mammogram abnormalities require biopsy, automated methods for extracting information from the biomedical literature, automated interpretation of images from x-ray crystallography, automated pharmacophore discovery to help guide drug design, and the use of computer technologies in home-care for patients with complex health problems. Collaborators include the UWCCC, McArdle, Computer Sciences, Pharmacy, Genetics, Biochemistry, GCRC.

**Image Analysis:** (Chung, Dyer). Dr. Chung and Dr. Dyer are developing algorithms for analyzing MRI images of the brain, where the algorithms perform segmentation, co-registration, and description. Collaborations exist with the Keck Lab (brain imaging), the Department of Ophthalmology and Visual Sciences, the Fundus Photograph Reading Center, Medical Physics, Mechanical Engineering, Waisman Center (speech kinematics), and the Department of Radiology (diagnostic and functional MRI).

## Methodological Research

Informatics faculty have independent research portfolios with peer-reviewed funding through NCI, NSF, DOD and other agencies. Some examples include:

**Link Discovery and Pattern Learning (Page, Shavlik):** Dr. Shavlik and Dr. Page have been involved in Defense Department-funded (DARPA, Air Force) methodological research focused on the development of multi-relational data-mining algorithms. This research has contributed to the growth of statistical relational learning (SRL), one of the most rapidly-growing areas of data mining and machine learning research (<http://www.dagstuhl.de/05051/>, <http://kdl.cs.umass.edu/events/srl2003/>).

**Advice-Taking Machine Learners (Shavlik):** There is much that human teachers can provide to learning machines than the traditionally used labeled examples and simple reinforcements. In another DARPA-funded project, Dr. Shavlik is developing robust algorithms that allow human teachers to provide advice, using simple English, to machine learners. The ability to instruct the software systems that one uses everyday, thereby personalizing software to match one's style of working, promises to have a revolutionary impact on how humans operate in our complex information-technology society.

**Biomedical Text Mining (Craven and Shavlik):** Dr. Craven and Dr. Shavlik have both been developing novel algorithms and systems for automatically (i) filtering biomedical articles for their

relevance to a particular information need, (ii) annotating the results of high-throughput experiments with literature-extracted keyphrases, and (iii) populating databases by extracting assertions from the literature.

**Machine Learning with Rich Data Sources and Interrelated Tasks (Craven):** In the context of developing methods for uncovering gene-regulatory elements and networks, Dr. Craven's group has devised a number of novel machine-learning algorithms that are applicable to problem domains that involve multiple data sources, sequence data, and/or interrelated learning tasks. Among the specific contributions of this research program are state-of-the-art algorithms for (i) refining the structure of stochastic context free grammars, (ii) taking advantage of relationships among multiple learning tasks to generate additional "weakly" labeled training examples from a pool of unlabeled examples, and (iii) representing and predicting elements in sequential data that overlap in arbitrary ways.

### **Clinical Informatics**

Members of the CISR continue the development of the Oncore Consortium, of which the UWCCC is the founding member. The goal of the consortium is to create software for an Oncology Collaborative Research Environment (Oncore) that provides integrated operational support of clinical research. This software continues to evolve through a process of continuous improvement and extension.

Last year, in addition to the UWCCC, the consortium members included the University of Iowa Holden Cancer Center, University of Minnesota Cancer Center, Vanderbilt-Ingram Cancer Center, Clinical Cancer Center at Stanford, Case Comprehensive Cancer Center, Cancer Institute of New Jersey, the UNC Lineberger Comprehensive Cancer Center and a local industrial partner, PercipEnz Technologies. During this year additional cancer centers have joined. These are the UNC Lineberger Comprehensive Cancer Center, Indiana University Cancer Center, Barbara Ann Karmanos Cancer Institute, University of Maryland Greenebaum Cancer Center, Simmons Cancer Center, and Milton S. Hershey Medical Center

Markey Cancer Center for a total membership of 15. All members now use common Oncore Technology for the conduct and management of their clinical trials and are beginning discussion of how to take advantage of this to engage in research projects involving the consortium membership in whole or part.

### **Computing Infrastructure**

The CISR provides computing support and infrastructure for UWCCC members and other shared services. The network, servers, storage and database facilities used on a daily basis by almost every UWCCC member have been upgraded with attention to capacity, reliability, and security. Distributed computing facilities for the execution of parallel programs now encompass over 130 machines and available online storage is now over five terabytes. The CISR is investigating ways to provide higher levels of availability especially for mission critical resources such as Oncore.

**Medical Informatics:** Informatics faculty have independent research portfolios with peer-reviewed funding through NCI, NSF, DOD and other agencies. Some examples include:

**Link Discovery and Pattern Learning (Page, Shavlik):** Dr. Shavlik and Dr. Page have been involved in Defense Department-funded (DARPA, Air Force) methodological research focused on the development of multi-relational data-mining algorithms. This research has contributed to the growth of statistical relational learning (SRL), one of the most rapidly-growing areas of data mining and machine learning research (<http://www.dagstuhl.de/05051/>, <http://kdl.cs.umass.edu/events/srl2003/>).

**Advice-Taking Machine Learners (Shavlik):** There is much that human teachers can provide to learning machines than the traditionally used labeled examples and simple reinforcements. In another DARPA-funded project, Dr. Shavlik is developing robust algorithms that allow human teachers to provide advice, using simple English, to machine learners. The ability to instruct the software systems that one uses everyday, thereby personalizing software to match one's style of working, promises to have a revolutionary impact on how humans operate in our complex information-technology society.

**Biomedical Text Mining (Craven and Shavlik):** Dr. Craven and Dr. Shavlik have both been developing novel algorithms and systems for automatically (i) filtering biomedical articles for their relevance to a particular information need, (ii) annotating the results of high-throughput experiments with literature-extracted keyphrases, and (iii) populating databases by extracting assertions from the literature.

**Machine Learning with Rich Data Sources and Interrelated Tasks (Craven):** In the context of developing methods for uncovering gene-regulatory elements and networks, Dr. Craven's group has devised a number of novel machine-learning algorithms that are applicable to problem domains that involve multiple data sources, sequence data, and/or interrelated learning tasks. Among the specific contributions of this research program are state-of-the-art algorithms for (i) refining the structure of stochastic context free grammars, (ii) taking advantage of relationships among multiple learning tasks to generate additional "weakly" labeled training examples from a pool of unlabeled examples, and (iii) representing and predicting elements in sequential data that overlap in arbitrary ways.



[webfeedback@biostat.wisc.edu](mailto:webfeedback@biostat.wisc.edu)

Copyright © 2006 The Board of Regents of the  
University of Wisconsin System