



DEPARTMENT PROFILE

Academic staff	22 FTE
Postdoctoral Research Assts	14.5 FTE
Undergraduates	110 students
M.Sc. in Applied Statistics	30 students
Postgraduate Research	55 students

Research Income (2007/08): £1.3m

RAE 2008 quality profile:				
4*	3*	2*	1*	FTE
40%	50%	10%	-	24.50

Oxford had the highest proportion (90%) of activity judged to be 'world leading' (4*) or 'internationally excellent' (3*), and the highest proportion of 4* activity in the UK.

TEACHING

The department leads the 3-yr BA and 4-yr M.Math. undergraduate courses in **Maths and Statistics**.



Graduate courses include a 1-year **M.Sc. in Applied Statistics**, and a part-time M.Sc. in Bioinformatics (jointly with Continuing Education, and currently under review), as well as the D.Phil. programme (which includes Doctoral Training Centre students researching bioinformatics and statistical genetics).

Graduate student funding is a major constraint: currently less than 50% of students are funded. The main sources are EPSRC, and department funds (Teaching Assistant Bursaries).

RESEARCH

The Department of Statistics is a centre for research in Statistical Science, including methodological developments and probabilistic

modelling, and their applications. It has distinguished individual researchers in addition to large and successful research groups in mathematical genetics and bioinformatics.

Individual research interests can be grouped into: **computer-related statistics, complex stochastic systems, applied probability, bioinformatics, statistical genetics, discrete mathematics and operational research**. A full list of research interests is included, below.

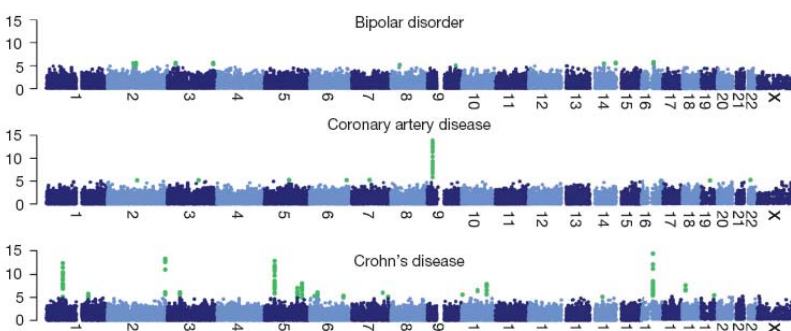
Since 2004, the department has been awarded over 60 research grants, with a total value of £9.6m. The main sponsors (either by number or size of awards) are: EPSRC, the Medical Research Council (MRC) and the Wellcome Trust.

Its research excellence is recognised through numerous honours – both individually (award of FRS to Donnelly, 2006) and collectively (RAE 2008 success).

As well as fundamental research in statistics and applied probability, there is much collaborative interdisciplinary research. Examples include:

Statistical Genetics

The largest and most successful study of the genetics of common human diseases, the Wellcome Trust Case Control Consortium, was led from the Department (**Donnelly**) with most of the methods development and analysis also in the Department. The Consortium's work, and subsequent collaborations, have been responsible for over 50 novel disease associations. The major paper won awards from *The Lancet* (Paper of the Year 2007), *Scientific American* (Research Leader of the Year 2007), *Nature* (Editor's Pick for 2007), and the *American Heart Association* (Top Paper for 2007). The Department also has a major role in the 1,000 Genomes Project (**McVean** co-chairs the analysis group), in addition to the development of novel statistical methods and software that is now widely used in genetics.



This image shows details of some regions where disease-causing genes were found from a genome-wide scan for 7 diseases (3 shown). The x-axis shows the 23 chromosomes shaded for clarity. Each point represents the strength of statistical evidence for a disease association at that position on the genome. Green points represent strong evidence.



Protein Interaction Networks

The figure shows the major component of the protein interaction network used for calculating the network statistics. The left-hand network is from *H. pylori* and the right-hand network is from *S. cerevisiae*. The colour of vertices indicates their degree (number of neighbours), where, for example, green is for proteins with degree less than 20, and yellow is for degree larger than 20.

Astrostatistics

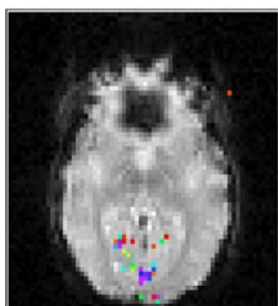
Recent and upcoming sky surveys scan the universe at an unprecedented rate, gathering data about millions of unknown objects. Detecting objects with a periodic light intensity allows us to identify and classify previously unknown pulsars. This statistical task is computationally challenging. We have developed a new algorithm that can detect pulsars at virtually the same efficiency as previously, yet using a factor of 10,000 less



computational resources. In collaboration with scientists from UC Berkeley and Cornell, this approach will be used in extensive searches for unknown objects in an upcoming satellite-based gamma-ray survey.

R project

The R project (www.r-project.org) is an international collaborative project that aims to provide a world-class environment for statistical computing, the development of new statistical methods and the graphical presentation of data. Prof. **Ripley's** group is one of the largest contributors to the project. Although originally developed for statistical research, R is now one of the most widely used of all statistical systems by researchers in many disciplines, and is close to universally used for teaching statistics at graduate level (and widely used for UGs). For just two examples, R has powered on-the-night election forecasting in the UK and in Austria, and much of the analysis of data from 'gene arrays' is done using R. It is also widely used in commercial settings, for instance in the pharmaceutical and financial industries, and for imaging.



One of the large-data applications for which R is used is **analysing brain images**. This figure shows a horizontal slice of a human brain, with colours indicating the response to a stimulus. The scale is in significance levels for activation in that square.

RESEARCH INTERESTS

Stochastic models, polygonal random fields, computer-intensive statistics (**Clifford**); Systems biology (**Dalby**); Protein structure evolution (**Deane**); Statistical Science (Chair), Statistical Genetics (**Donnelly**); Infinite dimensional stochastic processes and their applications (**Etheridge**); Probability distributions of genealogical processes (**Griffiths**); Probability and Combinatorics. (**Goldschmidt**); Bioinformatics (Chair), Evolution, population genetics, comparative genomics (**Hein**); Bayesian Statistics, Stochastic Simulation, Pattern Recognition (**Holmes**); Statistics (Chair), Forensic genetics, graphical models, causal inference (**Lauritzen**); Localisation, detection, characterization: genes that underlie human disease (**Marchini**); Probability statistical physics and theoretical computer science (**Martin**); Discrete mathematics, random structures and algorithms (**McDiarmid**) Population genetics, coalescent modelling, pathogen evolution and variation (**McVean**); Asymptotics for Lasso-type Estimators, Sparsity, Machine Learning (**Meinshausen**); Population genetics: stochastic models to understand variation (**Myers**); Bayesian inference, statistical methods, Monte Carlo, Applied Statistics (**Nicholls**); Network statistics and approximations in statistics (**Reinert**); Applied Statistics (Chair), Computer intensive statistics, pattern recognition, statistical software (the R project) (**Ripley**); Statistics in the Social Sciences (Chair), Methodology for social network analysis (**Snijders**); Biodemography, random dynamical systems, meta-analysis, human sex ratio (**Steinsaltz**); Levy processes, time changes, random trees and forests (**Winkel**)

FACILITIES

The department is centred on two (listed) Victorian villas in the Science Area, and also occupies space in the Peter Medawar Building and the Henry Wellcome Centre for Gene Function. A key priority for the department is to consolidate its activities on a single site.

The department has excellent computing facilities, including 2 student computer rooms.

EXTERNAL LINKS

- Risk Workshops
- Science Oxford
- BA Festival of Industrial Associations
- hosting Royal Statistical Society and related professional societies' events

KNOWLEDGE EXCHANGE

- entrepreneurial endeavour through registered patents and collaboration with companies and industries.
- In conjunction with collaborators from other departments, spinning out a company that applies local expertise in identifying genetic risk factors for common diseases.

For further information see: www.stats.ox.ac.uk