Statistics

SERUM ALBUMIN AS A ZINC CARRIER. Katarzyna B. Handing[1,2], I. G. Shabalin[1,2], M. Chruszcz[3] & W. Minor[1,2], [1]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22908, [2]New York Structural Genomics Research Consortium (NYSGRC), [3]Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208. Serum albumin (SA) is the most abundant protein in mammalian blood plasma (~600μM), making it the main determinant of plasma osmotic pressure. SA is able to bind and transport a wide variety of molecules including fatty acids, hormones, drugs and metal ions. SA is an important physiological transporter of several essential metal ions and assists in maintaining levels of free metal ions in blood. The aim of our study was to characterize zinc binding sites on SA in mammalian species. We have solved the crystal structures of human (HSA) and horse serum albumin (ESA) in complex with zinc. The presence of metals in the structures has been confirmed experimentally by X-ray anomalous scattering. Our data confirms that metal-binding site A is the primary binding site for Zn2+ and reveals the residues responsible for zinc coordination and the geometry of the binding sites. We see that site A is conserved between mammals. Additionally, pHs below physiological levels affect albumin affinity for zinc by protonation of zinc-coordinating histidine, which we show adopts different conformations at different pHs. We find several secondary zinc binding sites that are of biological interest, since site A may be disrupted by fatty acid and/or other metal binding. These findings expand our knowledge of metal transport and distribution in mammals. A better understanding of metal binding to SA is critical for the understanding of metal homeostasis and its contribution to disease states.

EXPLAINING THE TEN-YEAR INCREASING TREND IN KIDNEYS RECOVERED FOR TRANSPLANTATION BUT DISCARDED. Victoria. C. Garcia[1,2], B. J. Carrico[2] & D. E. Stewart[2], [1]Dept of Biostatistics, Virginia Commonwealth University, Richmond, VA 23198, [2]United Network for Organ Sharing, Richmond, VA 23198. Though the proportion of kidneys recovered for transplantation but discarded (discard rate, DR) have remained relatively stable from 2009 to present, the DR steadily increased from 1999 to 2009, from 13% to 19%. An analysis was performed to determine if this increase could be explained by a rise in the recovery of kidneys from a broader, more expansive donor pool, as opposed to more risk-aversive practices among transplant hospitals. Data on all deceased donors with a kidney recovered for transplantation from 1999 to 2009 were analyzed, with

donor characteristics collected by the United Network for Organ Sharing as the Organ Procurement and Transplant Network. From Oct. 25, 1999 to 2009, 65,292 deceased donors had 129,734 kidneys recovered for transplant, and 21,306 of them were discarded. After accounting for Kidney Donor Risk Index (KDRI) characteristics, the time trend was no longer statistically significant, suggesting changes in the characteristics of recovered kidneys explained the ten-year DR increase. Yet, also adjusting for biopsy-related factors reversed the predicted DR change over time, suggesting that, if the donor characteristics and the proportion of kidneys biopsied had remained stable, the DR would have actually improved slightly. However, after including the clinical decision of whether or not to pump a kidney, the residual time effect reversed again, suggesting that more pumping of kidneys may be a way to increase the available supply of kidneys for transplantation.

CHARACTERIZING THE VAGINAL MICROBIOME BASED ON A LARGE CROSS-SECTIONAL STUDY. Paul J. Brooks, David J. Edwards & Victoria V. Pokhilko, Department of Statistical Sciences and Operations Research, VA Commonwealth University. We conducted an analysis of 16S rRNA surveys of the vaginal microbiome based on samples from over 6,000 women. Subjects also provided medical, lifestyle, and demographic information. We investigate differences in microbiome composition by demographic factors, history of pregnancy complications, history of sexually-transmitted infections (STIs), the diagnosis at the time of sample collection, and pregnancy status. Vaginal microbiome profiles are typically dominated by a single bacterium, leading to a classification of samples into groups that we call *vagitypes*. Vagitype classifications facilitate the discovery of relationships between microbiome profile and clinical data. The presence or absence of Lactobacillus species and a diagnosis of bacterial vaginosis have been shown to play an important role in the reproductive health of a woman. Our analysis provides information about these patterns and suggests roles for other bacteria in health and dysbiosis.

ANALYZING HONG KONG AIR POLLUTION USING DIMENSION REDUCTION. H. Moradi Rekabdarkolaee, Edward L. Boone & Qin Wang, Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA, 23284. Dimension reduction and variable selection play important roles in high dimensional data analysis. Minimum Average Variance Estimation (MAVE) is an efficient approach among many others. However, because of using least squares criterion, MAVE is not robust to outliers or errors with heavy tailed distributions. In this paper, we propose a

robust extension of MAVE which can adapt to different error distributions. Our proposed estimate is shown to have the same convergence rate as the original MAVE. Furthermore, we combine the proposed method with adaptive LASSO to select informative variables. This new approach is illustrated through simulation studies and a data analysis on air quality of Hong Kong.

A DESIGNED EXPERIMENT APPROACH FOR THE TUNING OF OPTIMIZATION SOFTWARE. Toni P. Sorrell, J. Paul Brooks & David Edwards, Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA 23284. The tuning of optimization software is of key interested to researchers solving mixed integer programming (MIP) problems because the efficiency of the optimization software is can be greatly impacted by the solver's parameter settings and the structure of the MIP. Design of experiment methodology can provide insights into choosing good optimization solver's parameter settings for MIPs. A designed experiment approach is used to fit a model that would suggest settings of the parameters that provided the greatest impact on the primal integral, which is the comparison metric being used. Primarily, this research focuses on using classes of MIPs to not only obtain good parameter settings for a practitioner to use on future instances of the same class of MIPs, but to also gain understanding of why the settings work well for that class of MIPs. These experiments are conducted on three classes of MIP problems: survivable fixed telecommunication network designs, a formulation of Support Vector Machines (SVM) with Ramp Loss, and Coding Theory graphs. We compared our method to four other methods. Three methods come from the Selection Tool for Optimization Parameters (STOP), and they are random, greedy heuristic, and pairwise coverage. The last is CPLEX's and Gurobi's own tuning, which we consider a black box approach. An exhaustive search for the best parameter settings on a limited number of parameters provides a reference for the effectiveness of the methods used as a proof a concept for further experiments involving a large number of parameters.

DETECTING TINY SIGNALS IN MASSIVE DATA FROM HIGH-ENERGEY PHYSICS. Karen Kafadar, Department of Statistics, University of Virginia, Charlottesville, VA 22904. Experiments in high-energy physics provide terabytes of data, from which critical information about the state of matter, governed by the theory outlined in the "Standard Model", must be extracted. Opportunities abound for increased efficiencies in approaches to the data, from the design of experiments, to the collection of data, and finally to analysis and inference. Due to the massive amounts of data, from various sources

(different experiments from different collaborations, experiment-based simulations, etc.), new ways of analyzing the data to answer questions of interest are devised. This talk describes the framework for these experiments and illustrates methods for analyzing massive data sets from such experiments (with some mention of data sets from genomics and the Internet).