José Sánchez

AstraZeneca, Pepparedsleden 1, 431 50 Mölndal
+46(0) 76 768 3090
jose.sanchez1@astrazeneca.com

Profile

I am a statistician with a solid theoretical and applied background in mathematical statistics and data science. I have extensive experience in the statistical analysis of high-dimensional data (classical statistical learning methods, regularised regression and multivariate methods and machine learning methods) including clinical, laboratory and omics data. I have experience in design and analysis of pre-clinical and early (phase I and II) clinical data (selection and definition of endpoints, populations, power analysis, hypothesis testing, ANCOVA, mixed effects models, simulation, survival analysis), particularly in the cardiovascular, renal and metabolic areas. More recently I have applied Bayesian inference and machine learning methods for exploratory analyses.

The nature of my work at AstraZeneca is cross-functional and I collaborate with clinical experts, PK/PD scientists, biologists and chemists alike. For publications I have collaborated with scientists at Karolinska Institutet, Uppsala University, Sahlgrenska University Hospital, Fraunhofer-Chalmers Center, Lund University, University of Skövde and, internationally, with Cambridge University and Harvard Medical School.

Personal statement

Artificial Intelligence, AI, has become part of almost all areas of modern life ranging from home assistants like Alexa, to self-driving cars with Wikipedia article writing bots in between. AI advances itself in parallel to its applications, and drug development is not an exception in this race. Over the last 5 years, AstraZeneca has invested in talent and infrastructure to incorporate AI and its applications to drug development resulting in the creation of the Artificial Intelligence and Data Science department of which I am part of.

During my time in AstraZeneca I've kept current with the advance of AI and its applications to drug development. My research interests are in the development of AI methods for early clinical drug development, specifically, to incorporate machine learning methods as standard procedures for exploratory analyses in phase I and II clinical trials in order to identify target populations and inform the design of late phase clinical trials and personalised medicine.

Typically, early phase clinical trials collect data from a few hundred patients but for potentially thousands of variables including demographic, clinical, laboratory and omics data. I have done extensive research in the development of classical statistical learning methods in the low-sample high-dimensional setting; it is my intention now to continue with this work for neural networks and apply them in phase I and II clinical trial data.

As an interviewer for biostatistics positions at AstraZeneca I have seen candidates with statistical background struggle to put together relevant concepts they know about (such as randomised designs, hypothesis testing and regression modelling) as a whole in the context of a clinical studies design. This methods are part of basic statistical inference courses which, in order to be as general as possible and show a broad range of applications, may fail to help the students with interests in clinical trials.

For the last two years I have taught together with Ziad Taib two courses at Chalmers; Linear Mixed Effects Models and Design and Analysis of Clinical Trials. It is my intention to continue teaching this courses since they cover the required basic knowledge for the student interested in pursuing a biostatistics position in drug development, covering the statistical concepts named above (and more) and encompassing them in relevant contexts.

Although highly regulated by local and international health authorities, the design and analysis of clinical trials incorporates more and more machine learning, aiming not only for the analysis of high-dimension data but also for the design of more innovative, cheaper and faster design, including seamless phase II and III designs. I think this changes should be reflected in biostatistics

courses, so that they include state-of-the art AI applications in biology and medical research, particularly drug development. For this reason, I'll work together with the relevant people at the Mathematical Sciences department to restructure these courses to be more oriented to data science in drug development and attract a broader student audience.

I have been in charge of internal training at AstraZeneca and taught several lectures the University of Gothenburg and at Chalmers more recently. This has helped with my own training as a lecturer and kept me close to students, although I haven't been the main lecturer responsible for a whole university course since I was a Ph.D. student. Finally, I think that as an AstraZeneca employee and an adjunct lecturer at Chalmers i will be in a perfect position to continue with my research interests and strengthen the academic collaboration between the university and AstraZeneca.

Education

Ph.D. in Mathematical Statistics, December 2014. **Chalmers University of Technology.**

Thesis title: *Network models with applications to genomic data: generalization, validation and uncertainty assessment.* Supervisor: Rebecka Jörnsten.
We studied the problem of estimation of transcription networks, their validation and uncertainty. First, we extended graphical lasso methods in order to analyse different types of biological data across several groups. This method was then generalized to include estimation of expression levels, thus allowing to study its applications to tumour classification through discriminant analysis. We approached the validation problem by means mixture models applied to frequency statistics collected by bootstrap. The method selects the optimal network size by controlling the false positive rate and presents a robust final estimate that improves on naive bootstrap-threshold methods. Finally, we addressed the problem of uncertainty in estimation by extending the concept of point estimate and presenting instead a set of candidate networks. The candidate network estimators are based on bootstrap and information theory methods.
Together with biologists at Nelander Lab, Uppsala University, and Sahlgrenska Academy we developed a web-based tool, Cancer Landscapes `cancerlandscapes.org`, for the visualization and analysis of the estimated networks.

Lic. Phil. in Mathematical Statistics, May 2013. **Chalmers University of Technology.**

Thesis title: *Comparative Network Analysis of Human Cancer: sparse graphical models with modular constraints and sample size correction.*
We modelled transcription networks for several cancer classes and integrated different data types. We extended current methods in Gaussian graphical models taking into account biological constraints of modularity, sample size correction and instability of estimates via bootstrap. Joint work with biologists at Nelander Lab, Uppsala University, and Sahlgrenska Academy.

M.Sc. in Engineering Mathematics, May 2009. **Chalmers University of Technology.**

Thesis title: *Vehicle Damage Prediction from Advanced and Simple Systems Measurements.*
We designed and tested a regression model to predict brake damage by pressure for Volvo trucks. Joint work with Fraunhofer-Chalmers Centre and Volvo Trucks.

B.Sc. Honours in Applied Mathematics, 2000 to 2004. **Instituto Tecnológico Autónomo de México.**

Thesis title: *On residue theory and its applications.*

Training

Artificial Intelligence: Implications for Business Strategy, 2019-2010, **Massachusetts Institute of Technology.**

LEDA, leadership training, 2018, **Mgruppen.**

Professional Experience

AstraZeneca

Associate Director Biostatistics, Since October 2020.

Work as a project statistician in the Early Biometrics and Statistical Innovation group as per the principal scientist description. As an associate director I lead smaller teams to contribute in the delivery of the Early Biometrics and Statistical Innovation department's goals. I'm part of cross-functional teams tasked with the development and improvement of the company's guidelines, standard operating procedures and delivery models for our clinical trials. I have supervised a 6-month project as part of the industrial Ph.D. programme and I currently supervise a 2-year project as part of the Graduate Student programme.

Principal Scientist Biostatistics, July 2018 to September 2020.

Work as a project statistician in the Early Biometrics and Statistical Innovation group. As a project statistical I am responsible for the statistical contribution to design, decision-making, interpretation and communication of clinical studies within a project in early phase (phase I to IIb). I hold accountable the statistical staff/CRO so that all work is carried out with regards to AZ standards and external regulations and represent AZ and Statistics to Health Authorities for specific projects/indications. Together with the members of multidisciplinary clinical teams I develop design options and provide decision support to enable the business to make informed decisions, by applying expert skills and novel statistical approaches in both drug projects and exploratory studies. I support education in internally and externally by training in different statistical and machine learning topics.

Senior Scientist Biostatistics, December 2015 to June 2018.

Work as part of the multidisciplinary group of Quantitative Biology (QuBi), Discovery Sciences (DS). As a statistician at QuBi I was responsible for providing statistical support for the design, analysis and communication of results of preclinical in-vivo and in-vitro experiments, including the review of statistical analyses to enable informed investment decisions in the therapy areas portfolio. I collaborated closely with the bioinformaticians in the group to design and analyse RNA-Seq studies; during 2017 I was the main DS contact for RNA-Seq studies and was in charge of allocating the necessary QuBi resources for their analysis. As one of the most senior statisticians in the team I was in charge, together with the other most senior member, of managing the statistics capabilities, developing our interactions with the therapy areas, coaching new starters and participating actively in the recruitment process. I provided on-site training to in-vivo and in-vitro scientists about statistical design and analysis of experiments (yearly) and other topics (on-demand).

Bioinformatics Core Facility, University of Gothenburg

Statistician, February 2015 to December 2015.

Work as a statistician providing consultancy services for research groups in biological sciences at the Sahlgrenska Academy and the University of Gothenburg. Analysis of gene and protein expression, protein phosphorilation, microarray, qPCR, survival and survey data. Quality control for sequencing data, biological pathway enrichment analysis. Responsible for report generation and presentation of results to the clients.

Millward Brown Mexico, Mexico City, November 2003 to August 2007.

Senior Research Executive

Work in client service area. Wide variety of design, analysis and presentation of marketing research projects. Important projects with local and international clients: Bayer (pharma), BBVA Bancomer (banking), Nestlé (nutrition), Volkswagen (transport), Movistar Central America (telecommunications) and PepsiCo (soft drinks). Experience in work-on-site for PepsiCo.

My work consisted on applying the company's own methodologies (pre and post-test, brand equity, tracking and ad-hoc studies) as well as classical statistical analysis of data (survey sampling, parameter estimation, fitting of probability distributions, hypothesis testing). As a senior research executive I managed a 10-12 people team and was responsible for the generation and presentation of results to the clients.

Academic Experience

Publications

Quantification od apolipoprotein L1 in human plasma by LC-MS/MS and ELISA to monitor suppression in healthy volunteers. Bhat, M., Miliotis, T., Liying, L., **Sánchez, J.**, Carlsson, A-C., Althage, M., Davidsson, P., in preparation.

High Protein Diet Accelerates Diabetes and Kidney Disease in the BTBR ob/ob Mouse. Granqvist, A., Ericsson, A., **Sánchez, J.**, Tonelius, P., William-Olsson, L., Dahlqvist, U., Andersson, A.K., Tomic, T. Hudkins, K. Alpers, C., Pellegrini, G., Soderberg, M., American Journal of Physiology-Renal Physiology, 2020.

Discovery of Retinoic Acid Receptor Agonists as Proliferators of Cardiac Progenitor Cells Through a Phenotypic Screening Approach. Drowley, L., McPheat, J., Nordqvist, A., Peel, S., Karlsson, U., Martinsson, S., Müllers, E., Dellsén, A., Knight, S., Barrett, I., **Sánchez, J.**, Magnusson, B., Greber, B., Wang, Q., Plowright, A., Stem Cells Translational Medicine, 2019.

Barrier properties and transcriptome expression in human iPSC-derived models of the blood-brain barrier. Delsign, L., Dönnes, P., **Sánchez, J.**, Clausen, M., Voulgaris, D., Falk, A., Herland, A., Brolén, G., Zetterberg, H., Hicks, R., Synnergren, J., Stem Cell, 2018.

A Disintegrin and A Metalloproteinase-9 (ADAM9): A Novel Proteinase Culprit with Multifarious Contributions to COPD. Wang, X., Polverino, F., Rojas-Quintero, J., Zhang, D., **Sánchez, J.**, et al, American Journal of Respiratory and Critical Care Medicine, 2018.

Human iPS-Derived Astroglia from a Stable Neural Precursor State Show Improved Functionality Compared with Conventional Astrocytic Models. Lundin, A., Delsing, L., Clausen M., Ricchiuto, P., **Sánchez, J.**, Sarbish, A., Ding M., Synnergren, J., Zetterberg, H., Brolén, G., Hicks, R., Herland, A., Falk, A., Stem Cell Reports, 2018.

Hemostatic effects of the ticagrelor antidote MEDI2452 in pigs treated with ticagrelor on a background of aspirin. Pehrsson, S., Johansson, K.J., Janefeldt, A., Sandinge, A.S., Maqbool, S., Goodman, J., **Sánchez, J.**, Almquist, J., Gennemark, P., Nylander, S., Journal of Thrombosis and Haemostasis, 2017.

Efficient exploration of pan-cancer networks by generalized covariance selection and interactive web content. Kling, T., Johansson, P., **Sánchez, J.**, Marinescu, V., Jörnsten, R., Nelander, S., Nucleic Acids Research, 2015.

The Cancer Genome Atlas Pan-Cancer analysis project. Weinstein, J. et al. Nature Genetics. 2013. (Collaborator).

System-scale Network Modeling of Cancer Using EPoC. Abenius, T., Jörnsten, R., Kling, T., Schmidt, L., **Sánchez, J.**, Nelander, S. Advances in Systems Biology, Springer, 2011.

Network models with applications to genomic data: generalization, validation and uncertainty assessment. Ph.D. thesis. Chalmers University of Technology, 2014.

Comparative Network Analysis of Human Cancer: sparse graphical models with modular constraints and sample size correction. Licentiate thesis. Chalmers University of Technology, 2013.

Vehicle Damage Prediction from Advanced and Simple Systems Measurements. Master thesis. Chalmers University of Technology, 2009.

Supervision

Industrial Doctoral Project, October 2019 to April 2020, **Cambridge University and AstraZeneca.** In this 6-month project we investigated the relative importance of thousands of transcriptomics and proteomics biomarkers as predictors of diabetes. The data was from a relatively small exploratory study, falling in the low-sample high dimensional setting. We used random forests for binary response (diabetic and healthy subjects) and lasso regression for continuous response (HbA1c levels) with cross-validation to determine the most important predictors. Future work will include using these biomarkers for group classification in order to find treatment responders, which is key in better clinical trial

design and personalised medicine development.

Teaching

Design and Analysis of Clinical Trials, Chalmers, 2020.
Linear Mixed Effects Models, Chalmers, 2019.
Introduction to Machine Learning in R, AztraZeneca, 2019.
Good Statistical Practice in preclinical in-vivo studies, AstraZeneca; 2016, 2017.
Linear Statistical Models Design for RNA-Seq studies, AstraZeneca, 2017.
Multivariate Statistics (Regularisation methods module), University of Gothenburg, 2016.
Genomic data visualisation with Circos, University of Gothenburg, 2016.
Genomics and Bioinformatics (R module), University of Gothenburg, 2015.
Linear Statistical Models, Chalmers, 2013-2014.
Matematisk statistik och signalbehandling, Chalmers, 2011-2013.

Conference talks

Can I trust my network? Assessing network estimation uncertainty using local component resolution. Centre for Applied Biostatistics, University of Gothenburg, April 2015.
Joint estimation of transcription networks with applications to discriminant analysis. Linstat, Linköping, August 2014; Nordstat, Turku, June 2014.
Joint estimation of modular gene networks. Probability World Congress, Istanbul, July 2012; Nordstat, Umeå, June 2012.
Dynamic clustering of high-dimensional biological data. IMS, Gothenburg, August 2010.

Stipends awarded

Wilhelm och Martina Lundgrens Vetenskapsfond, 2015.
Stiftelsen GS Magnusons fond, Kungl. Vetenskapsakademien, 2012 and 2013.

Others

Swedish representative for the *European Federation of Statisticians in the Pharmaceutical Industry (EFSPI).*
EFSPI representative for the *Swedish Society for Medical Statistics (FMS).*
Review of scientific papers for *Journal of Statistical Software* and *Bioinformatics.*
Member of local organising committee for the *European Conference on Mathematical and Theoretical Biology*, Gothenburg, 2014.

Informatics Skills

Advanced programming in R and MATLAB with regular integration with C/C++ code.
Bayesian statistics with Stan.
Experience with Python and neural networks with Tensorflow and Keras.
Experience with Perl and Circos plots.
Ingenuity Pathway Analysis.
Bioinformatics tools such as Cytoscape, Integrative Genomics Viewer, online databases.
Common software for Windows and Unix platforms.
Parallel computing with computer clusters.

Languages

Spanish: native language.
English: fluent.
Swedish: fluent.

References

Marcela Dávila, Head of Bioinformatics Core Facility, GU. +46 31 786 97 17.
Lars Carlsson, Head of AI, Stena Line. +46 707 947 333