First Rousseeuw Prize awarded to Causal Inference

The King Baudouin Foundation has chosen the recipients of the 2022 biennial Rousseeuw prize for Statistics. This new scientific prize of 1 million dollars was created by Peter Rousseeuw, professor of statistics at KU Leuven, Belgium. The goal of the prize is to reward excellent statistical research with an important impact. Half of the prize amount will go to James Robins (Harvard University), and the other half jointly to Miguel Hernán (Harvard University), Thomas Richardson (University of Washington), Andrea Rotnitzky (Universidad Torcuato di Tella, Argentina) and Eric Tchetgen Tchetgen (University of Pennsylvania), for their groundbreaking methodological contributions to Causal Inference with applications in Medicine and Public Health. The latter four laureates were either trained or deeply influenced by Robins. They remain his main collaborators to this day.

The work honored by this award finds its origin in Robins' 1986 seminal paper on the estimation, from observational data, of the causal effects of treatments that vary over time. Before Robins' theory, there was little understanding, in the context of time-varying treatments, of the assumptions under which the estimates computed under a particular analytic approach had a causal interpretation, of the minimal assumptions needed to identify causal effects, and of the analytic methods that can be used to estimate effects. In fact, work in this area was essentially non-existent prior to Robins' article, even though time-varying treatments are ubiquitous in medicine and many other fields. In his paper, Robins formulated a new counterfactual causal model and associated identification theory that allowed for time-varying treatments and confounders, direct and indirect effects, and the feedback from one cause on another. This work resolved the long-standing problem of how to adjust for variables that are simultaneously intermediate on the causal pathway between past treatments and outcome, and confounders for future treatment. Robins' g-formula proved to be the keystone for resolving this problem under the assumption of no unmeasured time-varying confounders.

The g-formula was but the start of a long list of methodological innovations by the laureates, including methods for mediation analysis, optimal treatment regime estimation and sensitivity analysis, that helped launch a causal revolution in Statistics. Several of these innovations were motivated by concerns that estimation of the g-formula using parametric models could falsely find an effect when none existed, even in a randomized trial. To address this problem, Robins developed novel semiparametric causal models – Structural Nested Models and Marginal Structural Models – whose parametric component quantifies the causal effects of treatment over time.

In observational studies, even estimates of structural model parameters could be biased under the null hypothesis of no treatment effects. This is because of their reliance on the estimation of sets of high-dimensional time-specific nuisance functions: either the set of conditional treatment probabilities or the set of conditional expectations of the identified counterfactual outcomes of interest. To alleviate this dependence, Robins and Rotnitzky developed a pioneering theory of doubly-protected estimators. These are estimators of low-dimensional functionals whose construction requires estimating two high-dimensional, possibly vector valued, nuisance functions, but whose consistency only requires at least one nuisance estimator to be consistent, without knowledge of which one it is. In contrast to earlier estimators, doubly-protected

estimators are root-n consistent under much weaker constraints on the complexity (i.e., smoothness or sparsity) of the nuisance functions. As a consequence, doubly-protected methods have become widely adopted by epidemiologists, economists, and computer scientists, as well as by high tech companies such as Google, Amazon and Facebook. The theory of double protection grew out of, and extended, Robins and Rotnitzky's 1994 theory of semiparametric efficiency in coarsened at random models.

If the complexities of the unknown nuisance functions are too high, then even a doubly-protected estimator may have a bias that exceeds its standard error. To address this problem Robins, Tchetgen Tchetgen and colleagues developed a pathbreaking theory of point and interval estimation based on higher order U-statistics and scores. They applied their theory to construct rate minimax estimators of causal parameters that outperformed doubly-protected estimators in both confidence interval coverage and mean square error.

In causal inference, two different paradigms are used: causal directed acyclic graphs (DAGs), primarily used in computer science, sociology and philosophy, and counterfactual causal models, widely used in statistics and economics. Since only factual variables appear on causal DAGs, translation of the causal assumptions encoded by each of the two paradigms had not been straightforward. The translation problem has now been solved by Robins and Richardson's development of Single World Intervention Graphs (SWIGs). These graphs provide a simple unification of the two paradigms. SWIGs have accelerated the causal revolution by facilitating cross-disciplinary communication.

Outside of their collaborations with Robins, the laureates Rotnitzky, Tchetgen Tchetgen and Richardson have led many important developments in causal methods including the following. Rotnitzky's recent work on optimal adjustment sets is the first to bring together the theories of semiparametric efficiency and graphical models. Tchetgen Tchetgen's recent groundbreaking work on proximal causal inference has shown that negative outcome and treatment controls can be used to fully correct for bias due to unmeasured confounding; he has also derived a host of new conditions for the identification of treatment effects with instrumental variables. Richardson has pioneered models and algorithms for discovering causal structure from data, including the recent work on parameterization of the nested Markov model.

Through his expository papers of causal techniques and his causal analysis of numerous significant observational studies, Hernán has been instrumental in the dissemination of causal methods in the epidemiological community. He has recently established the CAUSALab, a Center at Harvard that uses data and causal methods to help decision makers make better decisions in medicine, public health, and policy.

The laureates' work has had both scientific and societal impact. It has been highly influential on statistical practice in medicine, epidemiology, and public health. Robins and Hernán proposed novel analyses of observational data based on the idea of emulating a hypothetical "target" randomized trial of the effects of time-varying treatments. The target trial framework has helped resolve high-profile disagreements between the results of earlier observational analyses and subsequent randomized trials in studies of the effect of post-menopausal hormone therapy on coronary heart disease, and of statin therapy on cancer. Their work showed that, in these cases,

disagreements historically attributed to lack of randomization in the observational studies could likely be attributed to the use of older, inappropriate statistical analyses. More recently, Hernán and colleagues applied the trial emulation approach to analyze, prior to the availability of clinical trial evidence, important observational studies of the timing of initiation of antiretroviral treatment in people with HIV, of the screening schedules for colorectal cancer and of the effectiveness and safety of Covid-19 vaccines. The latter has been instrumental to guide vaccination policies around the world. In all three cases, the findings were in agreement with the results of later clinical trials.

In closing, we refer the interested reader to the 2014 special issue of the journal *Statistical Science* on Semiparametrics and Causal Inference, which consists of a collection of twelve research papers, written by leading experts, that highlight the influence of Robins on their own research.

The international jury appointed by the King Baudouin Foundation selected the winners from the nominations received after a widely advertised call earlier this year. The jury consisted of its chair David Hand (Imperial College), Lutgarde Buydens (Radboud University Nijmegen), Probal Chaudhuri (Indian Statistical Institute), Dianne Cook (Monash University), Roger Koenker (University of Illinois), Yanyuan Ma (Penn State), David Scott (Rice University), David Steinberg (Tel Aviv University), Jane-Ling Wang (UC Davis), and James Zidek (University of British Columbia). For more information on the prize see www.rousseeuwprize.org .

The prize will be awarded in a ceremony taking place at KU Leuven, Belgium on Wednesday October 12, 2022.