Stat Biosci (2014) 6:163–165
DOI 10.1007/s12561-014-9121-5

EDITORIAL

Editorial

Robin Henderson · Vanessa Didelez · William Havercroft

Received: 14 October 2014 / Accepted: 20 October 2014 / Published online: 4 November 2014

The papers in this special issue are all associated with an invited workshop *Time for Causality: Causal Inference and Dynamic Decisions in Longitudinal Studies*, held in Bristol, UK, in April 2012 and sponsored by the 'SuSTaIn' programme (Statistics underpinning Science, Technology and Industry, http://www.sustain.bris.ac.uk/). The workshop brought together nearly 60 researchers interested in statistical methodology for causal inference and decision making from time-dependent data and modelling of dynamic systems. As well as mainstream statisticians, participants with backgrounds in machine learning, engineering control theory and a variety of application areas attended the workshop and participated in fruitful discussions.

The generic topic for the workshop can be described as follows. At time t, an input U_t is determined and a stochastic output Y_{t+} observed. The input may be (partially) directly controllable by an experimenter but the output Y_{t+} is not. Given the history or filtration $\mathcal{F}_{t-}$ of all previous inputs, outputs and other information leading up to time t, the purpose is often to choose U_t so as to achieve some objective measured in terms of future outputs $\{Y_{t+u} : 0 < u < \tau\}$, where the horizon τ may be finite or infinite. In other circumstances, the objective might be to determine the causal effects of changes in U_t, which can be hampered by the presence not only of sampling variability but also of time-varying confounding, meaning specialised methods are called for. There are also applications where possible interventions in the system are not specific and researchers are interested in modelling and describing a dynamic causal network, for example between events in cellular reaction systems. In most biostatistical applications, however, the input will often be a medical treatment, the

R. Henderson (✉)
Newcastle, UK
e-mail: Robin.Henderson@ncl.ac.uk

V. Didelez · W. Havercroft
Bristol, UK

output will be a measure of health, data will be available in the form of short sequences of observations on many subjects, and there will be no feasible opportunity to collect repeat data. By contrast, in typical engineering applications, usually a single subject is under study but it is closely monitored, with frequent observations, e.g. for an aircraft, updates to U_t will lead to movement of the ailerons, elevators and fin, in order to manipulate the attitude of the aircraft Y_{t+} during its flight mission.

Topics covered during the workshop included Q- and A-learning for dynamic treatments, control and sequential decision making in engineering, reinforcement learning, g-estimation, technical issues in causal inference, longitudinal and survival data analyses and stochastic kinetic process modelling. Applications included the relationship between mental health and alcohol consumption (nicely timed before the workshop drinks reception), asthma incidence prediction, causality for economic indicators and fault detection. The papers in this special issue reflect the balance of the workshop, covering underpinning theory, links with event history methodology, a detailed case study, dynamic treatment determination and control theory.

Dawid and Constantinou provide underpinning theory through careful discussion of conditions under which the effect of a decision strategy can be identified from data. Their paper provides, inter alia, a definition of a property they term simple stability, which is close in spirit to a formal and general version of the familiar no unmeasured confounders concept. Dawid and Constantinou then explore conditions under which this property might hold and inference from observational data be valid. They pay close attention to events that might have positive probability under one decision strategy but zero probability under another, and provide some intriguing counterexamples that illustrate how things can go wrong if there is insufficient attention to detail.

Arjas also provides underpinning theory and argues that we should take a stochastic process or event history interpretation for causal inference, with events in time treated as marked point processes, always conditional upon the past to date. His secondary argument is in favour of Bayesian inference and Markov Chain Monte Carlo methods. His paper is developed in the context of an example of how type of day care influences the risk of acute middle ear infection. The paper by Gerster, Ernæs and Keiding is also based around an application, in their case a careful study into the causal effect of educational attainment on completed fertility. The paper includes particularly clear descriptions of various modern techniques in causal inference, including feedback and the use of marginal structural models for longitudinal data. Through the analysis of data on a cohort of Danish women, the authors convincingly demonstrate the presence and importance of feedback between education and fertility processes.

Three papers in the issue concentrate on dynamic treatment allocation. The essential problem is causal inference in the presence of time-varying confounders, and the specific task is to derive adaptive decision rules for medical treatments or other interventions based on subject-specific characteristics and individual biomarker trajectories. In many applications, there are few decision times, a low number of possible treatments and a finite follow-up period. Hence, the advantages of flexible and robust techniques are evident, and this is the theme of all three papers on the topic. Moodie, Dean and Sun use flexible generalised additive models instead of the more traditional linear ones in the Q-learning approach. Barrett, Henderson and Rosthøj develop a doubly robust estimating equation approach to estimating regret functions in A-learning, while Rosthøj, Henderson and Barrett consider robustness to one form of missing data. This latter is an area that could be fruitful for further research, given that missingness can have at least three effects. One effect is the traditional effect on estimation, leading to at best inefficiency and at worst severe bias unless there is careful attention to assumptions, which are usually untestable. The second is missing data in the history $\mathcal{F}_{t-}$ as invariably the optimal treatment at time t will depend upon what has happened previously. And third, the focus of the Rosthøj et al paper is missing opportunities to change strategies in the future: an aggressive treatment strategy predicated on the assumption that it can be corrected at the next scheduled visit time is clearly not robust to the possibility that the visit may be missed.

The remaining paper in the issue is Taylor and Aerts work on control for dynamic systems. Control theory is concerned with the mathematical analysis of causal dynamical systems. It is a huge research topic in mathematical analysis, operator theory and applied engineering, yet there have been rather few attempts to marry this body of work to that in statistical causal inference. Recently, there has been growing interest in the use of control in biomedical applications, which typically have greater stochastic uncertainty and weaker repeatability than found in classical engineering application areas. Although control theory has been connected to biological systems for decades, developments in sensor technology mean that it is now possible to measure variables such as the heart rate of animals on-line, as illustrated in the Taylor and Aerts paper. Thus, there is need and scope for use of modern statistical estimation and inference methodology alongside modern control methods, and for the use of ideas and concepts from control in more traditional statistical applications. We hope that the Bristol workshop and this special issue of *Statistics in Biosciences* will stimulate further cross-disciplinary research in causal inference for dynamic processes.