Invited sessions for ISCB42 (Lyon, 2021)
Session 4

Challenges and opportunities for learning from long term disease registers

Organizers : **Els Goetghebeur : Ghent University, BE**
Linda Sharples : LHSTM, London, UK

Katie Harron, UCL, London England
Data linkage for creating electronic birth cohorts: handling bias due to linkage error.

Ingeborg Waernbaum, Departmentof Statistics, Uppsala University, Sweden
Challenges and opportunities for learning from long term disease registers: -Causal inference

Elizabeth Stuart, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA
Methods for combining experimental and population data to estimate population average treatment effects.

Discussant: Arvit Sjolander, Karolinska, Stockholm, Sweden

Data linkage for creating electronic birth cohorts: handling bias due to linkage error

Katie Harron[1]

[1] UCL Great Ormond Street Institute of Child Health, UK

Data linkage is a valuable tool for creating datasets with which to understand long term trajectories of health and disease. Linkage can provide a low-cost, efficient means of collecting extensive and detailed data on interactions with health and other services. These data can be used to create population-level electronic cohorts that offer the ability to answer questions that require large sample sizes or detailed data on hard to reach populations, and to generate evidence with a high level of external validity and applicability for policy-making.

Lack of access to unique or accurate identifiers means that linkage of the same individual across different data sources or time can be challenging. Errors occurring during linkage (false-matches and missed-matches) disproportionally affect particular subgroups of individuals and can lead to substantial bias in results based on linked data.

This talk will first describe methods for creating electronic birth cohorts using data linkage. We will then explore the impact of linkage error, drawing on examples from the literature. We will demonstrate a range of methods for evaluating linkage quality and discuss how bias due to linkage error can be handled within analyses.

Challenges and opportunities for learning from long term disease registers:

- **Causal inference**

Ingeborg Waernbaum

[1] Department of Statistics, Uppsala University, Sweden

Disease incidence registers are longitudinal data bases giving opportunities to study the development of chronic diseases in response to various treatments. For time-to-event emerging long term disease registers have at least two important advantages. They let many events build up over time, thus generating a rich information basis. They uniquely allow to study long term effects of treatment choices. When populations and available treatments change over calendar time, however, unavoidable administrative censoring patterns bring new estimation challenges as well as new questions on transportability of results.

Here, we describe two Swedish examples, the Swedish Renal Registry, recording incident cases of End Stage Renal Disease since 1991 and the Swedish Childhood Diabetes Register, a population-based incidence register, recording incident cases of childhood onset diabetes mellitus (T1DM) since 1977. We describe challenges and opportunities when controlling for baseline confounding for a point-treatment (immediate kidney transplant vs. start with dialysis) when the causal estimand is the difference in average potential survival curves. For the Swedish Childhood Diabetes Register, we describe selection bias and bounds when applying multiple inclusion criteria using extensions of previous results of Smith and Vanderweele (2019) and Sjölander (2020).

Methods for combining experimental and population data to estimate population average treatment effects

Elizabeth A. Stuart[1]

[1] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, USA

With increasing attention being paid to the relevance of studies for real-world practice (such as in education, international development, and comparative effectiveness research), there is also growing interest in external validity and assessing whether the results seen in randomized trials would hold in target populations. While randomized trials yield unbiased estimates of the effects of interventions in the sample of individuals (or physician practices or hospitals) in the trial, they do not necessarily inform about what the effects would be in some other, potentially somewhat different, population. While there has been increasing discussion of this limitation of traditional trials, relatively little statistical work has been done developing methods to assess or enhance the external validity of randomized trial results. This talk will discuss design and analysis methods for combining experimental and population data to assess and increase external validity, as well as general issues that need to be considered when thinking about external validity. Implications for how future studies should be designed in order to enhance the ability to estimate population effects will also be discussed.