At the centre for pharmacoepidemiology (CPE), Karolinska Institutet, we conduct research in the field of pharmaceuticals, with a focus on use and side effects. We use individual data from the Swedish national registers, primarily the prescribed drug register and the patient register. For more information about CPE see https://ki.se/en/meds/centre-for-pharmacoepidemiology.

The prescribed drug register contains information on all prescriptions filled by an individual, containing, among other things, the substance, dispensing date and amount dispensed. However, information on prescribed dosage and duration of treatment is missing, a central methodological problem is to estimate the time periods during which an individual is exposed. Several other countries have the same problem and many different solutions have been published.

The patient register contains information on all diagnoses that an individual received in hospital or in specialized outpatient care with dates of admission and discharge (for hospitalizations). However, there is a lack of data from primary care, e.g. visit to a health care center. Some side effects, e.g. anaphylactic shock, can be difficult to identify in the patient register, then an algorithm is needed that creates a proxy for the outcome by combining information from several data sources. Also, some important covariates, e.g. obesity, smoking and alcohol dependence are missing and sometimes replaced with proxies. There are a limited number of publications on proxies.

We have the following ideas that we believe would be suitable for a master's thesis in mathematical statistics.

1. How good is the Cox model? For example: How are estimated associations between drug exposure and time to side effect affected if a survival model other than Cox regression is used? How much is the result affected by non-proportional hazards?

2. How are estimated associations between drug exposure and time to side effect affected if one strictly applies causal inference with a counterfactual approach? If the response could be measured both with and without treatment for each individual? See https://en.wikipedia.org/wiki/Causal_inference for a description of causal inference.

3. How to best calculate and present sample size and/or expected precision in the planning stage of register-based studies?

4. Compare different ways of measuring adherence to drug treatment with data from the prescribed drug register (dispensing date and quantity). Various proposals for compliance measures have been published, e.g. proportion of days covered.

5. When we use observational data to determine whether a drug is effective or safe, we usually use another similar drug as a comparison. But it is also possible to either use a so-called negative comparator, i.e. a drug that is not expected to have the effect or side effect we are interested in, or a group of individuals who are not treated at all. Choosing a comparison group is a central problem in pharmacoepidemiology. How can we be sure that the comparison group is really comparable? We would like to map and explore existing methods as well as find new methods to measure comparability between different exposure groups.

6. Drug use patterns describe how new and old drugs are used over time using changes in the two measures prevalence (all users) and incidence (new users). Analysis of drug use patterns is central to pharmacoepidemiology. Usually, usage patterns are described with simple methods, e.g. different graphs. We would like to investigate whether time trend analyzes e.g. joinpoint and APC (Age Period Cohort) regression models can be used to improve understanding of drug use patterns.

Contact: Caroline Öberg, acting Head of CPE, Caroline.Oberg@ki.se

Marie Linder, senior statistician, Marie.Linder@ki.se