UCL Network of Applied Statisticians in Health

Target Trial Emulation Short Course*
Practical 2 (Stata and R)

28 November 2022

Introduction

In this practical we will revisit the concepts discussed in the afternoon lectures.

We will use simulated data generated as described in Figure 1. The data depict a scenario where data from electronic records on 5,000 newly diagnosed diabetic patients are collated to study the effect of treatment A (say Metformin) on a continuous outcome (say HbA1c) measured at the end of follow-up (time 2). Not all patients who initiate treatment at start of follow-up (time 0) continue while others may start a time 1.

The observed data consist of:

- time varying treatment $A = (A_0, A_1)$
- time-varying confounder $L = (L_0, L_1)$.
- end-of-study outcome Y

The data

The data are saved in two versions: long and wide. The long format version is called `Practical2_contY_long.dta`; the wide format version is called `Practical2_contY_wide.dta`. Their descriptions are below. Note however that although U is used to generate the data, it is not included.

```
Contains data from Practical2_contY_long.dta
 Observations:        20,000
    Variables:             5                  26 Nov 2022 09:46
```

*Part of the work funded by MRC Methodology Grant: MR/R025215/1

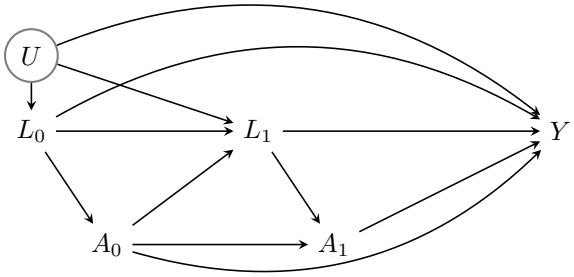

Figure 1: Directed Acyclic Graph (DAG) of a typical scenario for a time-varying exposure.

```
-------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------
id              float   %9.0g
t               byte    %10.0g
A               float   %9.0g                 Observed exposure at time t
Y               float   %9.0g                 Observed outcome at end of f-up
L               float   %9.0g                 Observed t-v confounder at time t
-------------------------------------------------------------------------
Sorted by: id  t


Contains data from Practical2_contY_wide.dta
 Observations:        10,000
    Variables:             6                  26 Nov 2022 09:42
-------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------
id              float   %9.0g
A0              float   %9.0g                 0 A
L0              float   %9.0g                 0 L
A1              float   %9.0g                 1 A
L1              float   %9.0g                 1 L
Y               float   %9.0g                 Observed outcome at end of f-up
-------------------------------------------------------------------------
Sorted by: id
```

Because we generated the data we also know the potential outcomes and hence also the true (observational-analog) ITT and PP. These are respectively: ITT=1.10 and PP=2.0.

Tasks

1. Examine the DAG: which arrow(s) would you remove if the data concerned an RCT?
2. Read and summarise the data using either data format (you choose!). If using R you could read the data using the `haven` package (see separate R code).
3. Examine the distribution of the outcome (remember that it is observed only at the end if follow-up. Note also that its value is repeated in each record (*i.e.* when t=0 and t=1) in the long format version).
4. How many patients initiate treatment at time 0? How many sustained treatment at time 1?
5. Estimate the conditional (confounder-adjusted) association between treatment initiation and the outcome using standard regression methods.
6. Estimate the observational-analog of the ITT effect of the treatment using IPW estimation of a marginal structural model (MSM). Follow these steps:
   (a) Specify the MSM you are targeting.
   (b) Use unstandardised weights to estimate the ITT.
   (c) Use standardised weights to estimate the ITT.
7. Estimate the observational-analog of the ITT effect of the treatment using g-computation. If using Stata you can use the `teffects` command (see separate Stata code).
8. Estimate the observational-analog of the PP effect of the treatment using IPW. This involves using the standardised weights of question 6 multiplied by the adherence weights described in the lecture (see separate Stata and R codes for guidance).
9. Interpret all results.