Causal Survival Analysis Workshop

Eleanor Murray, created with Ellen Caniglia
Lucia Petito

November 18, 2019

Contents

1 Overview

This workshop is designed to provide an overview to causal inference for survival outcomes with point exposures or time-varying exposures for static interventions (see Section 1.4) using inverse probability weighting. The basic concepts here also apply to other types of exposure strategies, although these may require additional design or analysis considerations.

This workshop describes the basic concepts needed to estimate causal effects on survival outcomes with applied examples. For a more detailed technical introduction to survival analysis and causal inference see Chapter 17 in Hernan & Robins 2019 [4].

The workshop is organized around three exercises:

1. Drawing your directed acyclic graph
2. Intention-to-treat effects
3. Per-protocol effects using inverse probability weighting for treatment-confounder feedback

1.1 Workshop materials

This workshop uses data created by simulating the covariate structure of the Coronary Drug Project randomized controlled trial[1]. The code for the workshop can be found at https://github.com/eleanormurray/NeuRA_Sydney_2019 and is available for `R`, `SAS`, and `Stata` software. Each software has a different preferred data file format, so be sure to download the correct version of the `trial1` data.

1.2 Background: The Coronary Drug Project trial

The US National Heart Blood and Lung Institute sponsored the Coronary Drug Project, a double-blind placebo-controlled randomized trial conducted between 1966-75, to determine the safety and efficacy of a set of drugs for secondary prevention of mortality among men with a history of myocardial infarction (heart attacks)[1].

The trial initially compared 5 active treatments, high- and low-dose equine estrogen, dextrothyroxine, nicotinic acid, and clofibrate, to placebo. However, three treatment arms (high- and low-dose estrogen, and dextrothyroxine) were stopped early for high risk of adverse events. We focus here on the comparison between **clofibrate** and **placebo** [3].

Clofibrate is a lipid-lowering agent first created in 1966 that works to increase lipoprotein lipase activity to decrease high cholesterol and triglyceride levels. In the U.S. it was initially sold under the trade name "Atromid-S", but in 2002, FDA approval was revoked due to adverse effects [9]. In the trial, patients were instructed to take 600mg 3x per day. Placebo here was sugar pills, designed to look like clofibrate, and taken on the same schedule.

Adherence to treatment in this trial was defined by the physician at each quarterly visit throughout follow-up, who visually inspect the bottle of pills to describe the adherence as "good" ($\geq$ 80% of pills used) versus "poor" (<80% of pills used).

Table 1 below describes relevant design components of the Coronary Drug Project protocol for this workshop.

1.3 Survival analysis basics

Often we are interested in understanding the average causal effect of an exposure not just on the occurrence of an outcome but on how long it takes for the outcome to occur (the **time to event**). The classic example for understanding survival analysis is estimating the effect of an exposure on

Table 1: Trial Description

Protocol	Description
Eligibility criteria	Men with a history of myocardial infarction in previous 3 months, 30-64 years old at baseline
Treatment arms	5 lipid-lowering active drugs vs placebo
Follow-up	Begins: at randomization Ends: earliest of 5 years after baseline, loss to follow-up, or death
Outcome	All-cause mortality within 5 years
Causal contrasts of interest	1) Intention-to-treat effect 2) Effect of good adherence to trial protocol versus poor adherence in the placebo arm 3) Per-protocol effect of continuous adherence to treatment versus placebo

the time to death, but we can conduct survival analyses for any event as long as the timing is available and of interest.

Survival analyses are complicated by the fact that we (almost) never have complete information on all individuals – even for an outcome like death which will eventually happen to everyone, we never know the time of death for all individuals. Instead, our data end at some fixed time after which we have no additional information on study participants. This is called **administrative censoring**.

In many longitudinal or follow-up studies, individuals are also **lost to follow-up**. This is another type of censoring and we typically have to decide what assumptions to make about who is lost to follow-up and why. One increasingly common approach is the use of **inverse probability of censoring weights**. Inverse probability of censoring weights are constructed similarly to inverse probability of adherence or treatment weights described in this workshop. In the current workshop, we have created a simulated data in which no loss to follow-up occurs, so we will not discuss loss to follow-up further.

1.4 Defining exposure strategies for survival analyses

Survival analyses can be conducted for point exposures or sustained exposures. We briefly describe these and considerations for defining exposure strategies for comparison.

A **point exposure** is an exposure that happens at a single time point. An example might be receiving one of two types of knee surgery, or being offered a single colorectal cancer screening sigmoidoscopy. Since point exposures happen only once, they typically require only control for **baseline confounding**. (However, estimating causal effects of point exposures on time to event may still require adjustment for time-varying confounders of loss to follow-up and the outcome.) Randomization is an example of a common point exposure.

A **sustained exposure** is an exposure that happens over time. An example of an sustained exposure is 'take aspirin once per day'. Since sustained exposures happen over time, they require adjustment for **time-varying confounding**. As we will see in Exercise 1, adjusting for time-varying confounding can *induce* bias if the confounders are affected by prior exposure. For this reason, whenever we have time-varying confounding we should use the **g-methods**: inverse probability weighting of marginal structural models, the parametric g-formula, or g-estimation of structural nested models.

Sustained exposures can be further categorized into static or dynamic exposure strategies. A **static exposure** is a sustained exposure where the value of exposure at any given time does not depend

on any time-varying individual characteristics. For example, 'continuous treatment' is an example of a sustained exposure. A **dynamic exposure** strategy is one where the value of exposure at any given time depends on time-varying individual characteristics, such as "take treatment until a side effect develops". Static and dynamic exposure strategies both require the use of g-methods, but dynamic exposures can require more complex stabilization of inverse probability weighting. In the current workshop we will focus on point exposures and static sustained exposure strategies.

Finally, for some exposures we may want to allow flexibility in when the exposure happens. We can do this by specifying **grace periods**. The use of grace periods adds an extra layer of complexity which we will not discuss in this workshop. The general approach requires copying or "cloning" all person-time during the grace period and censoring copies based on an individual's actual exposure history and our exposure strategy of interest. Adjustment for censoring of clones can be done using inverse probability weighting. A similar cloning and censoring approach can also be used for dynamic sustained exposures.

1.5 Randomized trials, pragmatic trials, and the target trial framework for observational data

Causal inference for survival analyses can be conducted using experiments, such as randomized clinical trials (RCTs) or pragmatic randomized trials, or observational studies. The main difference between causal survival analysis in RCTs or other experiments compared to observational studies is baseline confounding. RCTs typically do not need to worry about baseline confounding, because (assuming sufficiently large sample sizes) randomization generally guarantees no confounding between trial arms.

However, randomization is only protective against confounding at the time of randomization. For survival analyses, we are by definition interested in things that happen after randomization. Therefore, we may need to worry about post-randomization confounding for treatment adherence and/or loss to follow-up even in an RCT.

For causal survival analyses of observational studies, we also need to consider two additional challenges: identification of **time zero**, and **well-defined interventions**. In an RCT, we know when the study should start (i.e. time zero) because we assign exposure to everyone. However, in an observational study we don't know this because we don't assign exposure. Similarly, in an RCT because we have to intervene to give everyone the exposures of interest we are virtually guaranteed that the **consistency** assumption of causal inference holds. In an observational study, we may not be sure how people got the exposure status they actually had, so we may not be confident about consistency – using well-defined interventions to specify our exposure strategies of interest is a way to support the validity of this assumption.

One simple way to structure your study design and analysis to minimize the potential for bias in the choice of time zero and ambiguity in interpretation because of a lack of well-defined interventions is to use the **target trial framework**. In short, this means that we should think about a hypothetical randomized trial we would like to do and use this to design our observational study.

In this workshop, we use the example of a real randomized trial. However, everything we present here can be easily translated to the context of a hypothetical target trial and the corresponding observational data analysis.

1.6 Overview of the simulated workshop data

The Coronary Drug Project is real patient data, so sharing is restricted. Instead, we have created a simulated data set based on the relationships between a subset of the variables recorded in the Coronary Drug Project. The simulated data includes only two of the original six trial arms: placebo, and clofibrate.

The data are in **long format**, which means that there is one row, or observation, for each visit for each individual. Although the original trial had loss to follow-up, for simplicity we have chosen to simulate data with no loss to follow-up. Therefore, everyone is in the trial for the full five years, or until follow-up ends due to death. In your research you will likely have loss to follow-up and will need to make assumptions about the relationship between loss to follow-up, the exposure, the outcome, and measured and unmeasured covariates.

Table 2 is the **data dictionary** for the simulated data. It describes the available variables that you can use in your analyses, the notation we will use for them throughout this workshop, and the possible values of these variables. Note that for those covariates which are measured at each visit t the data also contains a variable with the baseline value. For example, `chf` is the variable name for coronary heart failure recorded at visit t, and `chf_b` is the corresponding variable for coronary heart failure recorded at baseline. At baseline (visit $t = 0$) the values of `chf` and `chf_b` will be identical for a given individual.

Table 2: Data Description

Variable	Notation	Definition	Values
simID	id	ID variable	Range: 0 to 4042
rand	Z	Randomization group	0: placebo; 1: clofibrate
visit	t	Visit number: visits occur quarterly	Range: 0 to 14
death	Y_t	Death recorded at visit t	0: alive; 1: died T $\in [t, t+1)$
adhpre0bin	A_{-1}	Adherence to placebo during pre-randomization run-in period	0: adherence $\geq$ 80%; 1: adherence $<$ 80%
adhr_b	A_0	Adherence to placebo at baseline (recorded at a special visit 2 weeks after randomization)	0: adherence $\geq$ 80%; 1: adherence $<$ 80%
adhr	A_t	Adherence to assigned treatment at visit t	0: adherence $\geq$ 80%; 1: adherence $<$ 80%
mi_bin	L	Myocardial infarction at baseline	0: 1 or 2; 1: $\geq$2
AntiHyp	L	Antihypotensive medication use at visit t	0: no; 1: yes
AnyQQS	L	ECG findings: Q/QS pattern at visit t	0: no; 1: yes
AnySTDep	L	ECG findings: ST depression at visit t	0: no; 1: yes
AP	L	Angina pectoris at visit t	0: no; 1: yes
CardioM	L	Cardiomegaly at visit t	0: no; 1: yes
CHF	L	Coronary heart failure at visit t	0: no; 1: yes
DIUR	L	Diuretic use at visit t	0: no; 1: yes
FVEB	L	ECG finding: Frequent ventricular beats at visit t	0: no; 1: yes
HiHeart	L	High heart rate at visit t	0: $<$70 bpm; 1: $\geq$ 70 bpm
HiSerChol	L	High serum cholesterol at visit t	0: $<$250; 1: $\geq$ 250
HiSerTrigly	L	High serum triglyesterol at visit t	0: $<$5.0; 1: $\geq$ 5.0
IC	L	Intermittent claudication at visit t	0: no; 1: yes
NIHA	L	New York Heart Association class at visit t	0: no limitations; 1: any limitation
OralHyp	L	Oral hypoglycemic agents use at visit t	0: no; 1: yes
VCD	L	ECG finding: Ventricular conduction defect at visit t	0: no; 1: yes

2 Directed acyclic graphs for survival analysis

2.1 Choosing a causal estimand

The first step in estimating causal effects on survival outcomes is to define the causal estimand of interest. In randomized trials, there are two classes of causal effect which we might want to estimate: the **intention-to-treat** effect, or **ITT**, and the **per-protocol effect**. In an observational study, we can typically only estimate the per-protocol effect.

The ITT is the effect of randomization to one trial arm versus the other. In the Coronary Drug Project trial, the intention-to-treat effect can be described as the difference in survival that would have been observed over 5 years if everyone had been randomized to clofibrate compared to if everyone had been randomized to placebo. The estimate of the intention-to-treat effect depends on the adherence patterns that actual occurred in the trial, so it can have generalizability problems if adherence patterns change.

The per-protocol effect is the effect of adherence to the treatment strategy defined by the study protocol. The interpretation of a per-protocol effect is therefore trial-specific, and more than one per-protocol effect definition is possible for a given trial. Per-protocol effects are often estimated in pragmatic trials because patients and providers want a measure of effectiveness that is not influence by adherence. Analyses of pragmatic trials can be especially challenging because they are susceptible to confounding and selection bias often associated with observational data analyses. The per-protocol effect is also typically the only effect that we can estimate from an observational study.

2.2 Exercise 1: drawing the DAG

Before we can estimate these causal effects, we need to think about the structure of relationships between variables in our trial. Using a directed acyclic graph (DAG) can be a helpful way to visualize the potential for bias due to confounding and selection bias [5].

We will work through the DAGs for the intention-to-treat effect and the per-protocol effect in a randomized trial. These DAGs will also be useful for thinking about target trials. When emulating a target trial with observational data, your DAG will look similar but will not have a node for randomization and instead may need additional baseline confounding nodes.

By convention, we use the following notation:

- Z indicates randomization assignment.
- A indicates treatment received. This can be further separated into A_0 for treatment received at baseline, and A_t for treatment received at a follow-up time point t. In addition, we can add an overbar to indicate treatment history up until a given time t: $\bar{A}_t$. Finally, we use lowercase a to represent an specific value of treatment received, and upper case A to represent the variable treatment received.
- L indicates the vector of measured covariates. We can specify L_0, L_t, $\bar{L}_t$ and l in the way as for A above.
- Y indicates the outcome, and we are typically interested in Y_t for survival analyses. We can use Y^a or Y^z to indicate the counterfactual outcome under either an intervention on treatment received or randomization status.
- U represents the vector of unmeasured covariates that we may be concerned about as potential threats to exchangeability.

2.2.1 DAG for the intention-to-treat effect

The intention-to-treat effect can generally only be estimated from a randomized trial, and is an estimate of the causal effect of random assignment. When everyone in the trial initiates treatment upon randomization, the intention-to-treat effect can further be interpreted as the effect of treatment initiation, and this effect can also be estimated from observational data.

For this exercise, consider the intention-to-treat effect as the effect of randomization. You can ignore loss to follow-up.

Question 1

In the space below, draw a DAG for the intention-to-treat effect in the Coronary Drug Project. In thinking about which variables to include in your DAG, consider whether A is necessary for estimating the intention-to-treat effect.

2.2.2 DAG for the per-protocol effect

The per-protocol effect is the effect of receiving treatment according to the trial protocol. This effect can be estimated in randomized trials and is typically the effect of interest in observational studies, especially when we want to know about the effect of a sustained exposure. Per-protocol effects are also patient-centered causal effects, and may be important for shared decision-making [6].

Question 1

In the space below, draw a DAG representing the per-protocol effect in the Coronary Drug Project. Remember that the Coronary Drug Project assigned participants to sustained use of clofibrate or placebo over 5 years. When drawing a DAG for multiple time-points, we commonly simplify by drawing only baseline and one follow-up time. In choosing variables to include on your DAG, think first about those variables which are required for the definition of the effect and then think about what measured and unmeasured variables might be common causes of each pair of variables on the graph.

3 Exercise 2: Estimating intention-to-treat effects

Now that we've described our assumptions about the structure of the causal relationships between variables in our trial, we will estimate the intention-to-treat effect. The intention-to-treat effect is the effect of randomization on the outcome, `death`, regardless of whether or not individuals adhered to the treatment protocol they were assigned to.

We can describe the intention-to-treat effect of randomization to clofibrate versus placebo on 5-year mortality using a variety of effect measures.

In this exercise we will estimate three different intention-to-treat effect measures using different approaches and compare and contrast the results.

The cumulative incidence difference by 5 years

$$\text{RD} = \Pr[Y^{z=1} = 1] - \Pr[Y^{z=0} = 1]$$

The cumulative incidence ratio by 5 years

$$\text{RR} = \frac{\Pr[Y^{z=1} = 1]}{\Pr[Y^{z=0} = 1]}$$

The hazard ratio over 5 years (where λ^z is the counterfactual hazard)

$$\text{HR} = \frac{\lambda^{z=1}}{\lambda^{z=0}}$$
$$\lambda^z = \Pr[T^z = t | T > k]$$

3.1 Data exploration

Before we estimate the intention-to-treat effect, let's get familiar with the data. Make sure you have the workshop materials downloaded and unzipped. In your preferred software, open the code:

- For R, double click on the '.Rproj' file in the 'NeuRA_Sydney_2019' folder to open the workspace.
- For SAS, navigate to the 'SAS' folder and double click the '.sas' file.
- For Stata, navigate to the 'Stata' folder and double click the '.do' file to set Stata's working directory to the workshop folder.

Run the code in **Code Section 0**. This ensures that all required libraries are installed (for R), the working directory is correctly assigned to the workshop folder (R & SAS), and the data is loaded into the software (R, SAS, Stata).

Once you have your coding environment set up, run the code in **Code Section 1** and answer the following questions:

Question 1

How many person-visits are in this dataset?

Question 2

How many individuals are in this dataset?

Question 3

How many individuals had at least 10 visits?

Question 4

In the example trial, individuals could be randomized to either placebo (`rand = 0`) or active treatment (`rand = 1`). The variable `rand` is constant within individuals across all visits. How many individuals were randomized to each trial arm?

Question 5

How many individuals died overall? How many individuals died in each treatment arm? What was the cumulative probability of mortality by 14 visits in each treatment arm?

3.2 Kaplan-Meier survival curves

Now that we have a sense of our data, we will use the Kaplan-Meier method to obtain a **non-parametric** estimate of survival separately in each trial arm. Non-parametric estimates don't require us to make any modeling assumptions about functional forms or variable relationships, beyond our causal inference assumptions. Since we don't have any baseline confounding, this will give us an estimate of the counterfactual survival if everyone had been assigned to clofibrate and if everyone had been assigned to placebo.

Run the code in **Code Section 2** to create the non-parametric survival probability estimates for each visit and plot the estimated survival curves. Then answer the following questions:

Question 1

Sketch your survival curve in the space below.

Question 2

Look at the summary table. How does the estimate of survival at visit 14 compare to the proportion of individuals who died in each arm separately that we calculated in the previous section?

Question 3

Based on the graph you created and the summary of the Kaplan-Meier fit, do you think there is a difference in survival at 14 visits between the treated and placebo arms? Is this a statistically significant difference? A clinically relevant difference?

3.3 Using models to estimate the intention-to-treat effect

In this section, we will use two different modeling approaches to estimate the intention-to-treat effect: **Cox proportional hazards** regression and **pooled logistic** regression.

Cox proportional hazards regression is a **semi-parametric** modeling approach – we make the assumption that the hazard ratio is constant over follow-up and in return we do not need to make any assumptions about the functional form of the hazards. If you only want to estimate the hazard ratio, Cox proportional hazards regression makes the fewest assumptions.

Pooled logistic regression is a **parametric** modeling approach – we need to make assumptions about the functional form of the baseline hazard and about whether the hazard ratio is constant or time-varying. In return, we can use the results of this model to estimate not only the hazard ratio, but also the survival, cumulative incidence (risk) difference, and risk ratio. The pooled logistic regression model estimates the discrete time hazard, so one important caveat is that we need the outcome to be rare in each time interval.

3.3.1 Unadjusted intention-to-treat effects

To get familiar with how these two modeling approaches work, we will start by estimating the unadjusted hazard ratio using both models.

$$\text{Cox proportional hazards: } \lambda(t \mid Z) = \lambda_0(t)\exp(\alpha_1 Z)$$

To estimate the hazard ratio from a Cox model, we only need to know the amount of follow-up time (`maxVisit`) and vital status at the end of follow-up (`deathOverall`) for each individual. In `R` and `SAS`, we can use our new `baseline` dataset; in `Stata`, we restricted to visit 0 when we used `stset`.

For Cox models, we need to specify how to handle ties (individuals with the same event time). There are a variety of options. We use the Breslow method here, which is a common software default.

$$\text{Pooled logistic regression: } logit[\Pr(Y_t = 1 \mid Z)] = \beta_{0,t} + \beta_1 Z = \beta_0^* + \beta_1 Z + \beta_2 t + \beta_3 t^2$$

To estimate the hazard ratio from a pooled logistic regression model, we need to use the full `trials` dataset and we need to include time as a covariate in our model to specify a functional form for the baseline hazard. We want to allow time to be included in a very flexible manner (polynomials, splines, categorical). In our trial, the time scale was `visit`, which numbered from 0 to 14 and occurred every 3 months. In this workshop, we have chosen to include `visit` as a quadratic polynomial for simplicity. In your research, you may find splines give a better fit.

Since pooled logistic regression uses multiple observations per person, we need to correct our standard errors. One option is to use **robust** ('sandwich') standard errors, which give us valid, but conservative confidence intervals that will be slightly wider than necessary (so the interval will fall around the true value more than 95% of the time). A better option is to use bootstrapping to get valid, non-conservative confidence intervals.

Run the code in **Code Section 3a**, and fill out the first 3 rows of Table **??**.

Exercise 2 results

		Cox proportional hazards	Pooled logistic regression
Unadjusted	Coef		
	SE		
	HR		
Adjusted	Coef		
	SE		
	HR		

Question 1

Do the results from the *unadjusted* Cox proportional hazards model and the *unadjusted* pooled logistic regression model match?

Question 2

What is the causal interpretation of these unadjusted hazard ratios?

Question 3

What assumptions are we making to give this hazard ratio a causal interpretation?

3.3.2 Baseline-adjusted intention-to-treat effects

Next, we will calculate adjusted intention-to-treat hazard ratios using both Cox proportional hazards and pooled logistic regression. In randomized trials, it is common to adjust the intention-to-treat analysis for known prognostic factors for the outcome. This can reduce the variance and protect against random confounding. We will adjust for all the baseline measurements of the covariates listed in Table 2. Baseline covariates all have a suffix of `_b` to denote they were measured at baseline.

The models we will be fitting are as follows.

Cox proportional hazards:

$$\lambda(t \mid Z, L_0) = \lambda_0(t)\exp(\alpha_1 Z + \alpha_2 L_0)$$

Pooled logistic regression:

$$logit[\Pr(Y_t = 1 \mid Z, L_0)] = \beta_{0,t} + \beta_1 Z + \beta_2 L_0 = \beta_0^* + \beta_1 Z + \beta_2 L_0 + \beta_3 t + \beta_4 t^2$$

Run the code in **Code Section 3b**, complete the last 3 rows of Table **??**, and answer the following questions.

Question 1

Do the results from the *adjusted* Cox proportional hazards model and the *adjusted* pooled logistic regression model match?

Question 2

Do the results from the *adjusted* models match the results from the *unadjusted* models?

Question 3

What is the causal interpretation of these adjusted hazard ratios?

Question 4

What assumptions are we making to give this hazard ratio a causal interpretation?

Question 5

Are these assumptions different from the assumptions required for the unadjusted hazard ratio?

3.4 Standardizing over baseline covariates to estimate marginal causal effects

Sometimes we may want to adjust for baseline covariates but retain our interpretation as the average causal effect without making additional assumptions about effect heterogeneity.

In randomized trials, adjusting for baseline predictors of the outcome can decrease our variance, and adjusting for baseline predictors of the outcome that also happen to be unbalanced between treatment arms can remove random confounding (see Chapter 10 in [4] for details on random confounding). In observational studies, adjusting for baseline covariates that are predictive of the outcome and of exposure is important for removing baseline confounding.

One option for estimating the average (or marginal) causal effect with adjustment for baseline confounders is to **standardize** our conditional estimate across baseline covariates. For survival outcomes, standardization requires an estimate of the baseline hazard, and therefore we cannot use a Cox proportional hazards model to obtain the average causal effect.

Instead, we use a pooled logistic regression to calculate average counterfactual survival and risk, and the average hazard ratio and cumulative incidence ratio. When using pooled logistic regression to obtain estimates of survival, we need to provide our model with some additional flexibility in the functional form of the baseline hazard. We will do this by including interaction terms between randomization arm and our function of time.

We will calculate counterfactual survival curves (one for survival if everyone was assigned to treatment, one for survival if everyone was assigned to placebo) and use these to estimate the average hazard ratio, the average cumulative incidence ratio, and the average risk difference at 14 visits, if everyone had been assigned to clofibrate vs placebo.

To do this, the **Code Section 4** follows 6 steps:

(1) **Data processing.** In R and SAS, we need to start by adding the product terms `randvisit` and `randvisit2` to the `trial` dataset. In Stata, we can skip this step.

(2) **Pooled logistic regression with interaction terms.** Estimate the parameters of a pooled logistic regression model including interaction terms created above. The code below is exactly the same as the adjusted pooled logistic regression model we ran in Code Section 3, except that we have added our interaction terms.

$$logit[\Pr(Y_t = 1 \mid Z, L_0)] = \beta_{0,t} + \beta_1 Z + \beta_2 L_0 + \beta_3 Zt + \beta_4 Zt^2$$

$$= \beta_0^* + \beta_1 Z + \beta_2 L_0 + \beta_3 Zt + \beta_4 t + \beta_5 t^2 + \beta_6 Zt^2$$

(3) **Simulated data: treated.** Create a new dataset which includes 1 copy of every person at baseline, and force everyone in this dataset to be assigned to treatment (don't forget to recreate the interaction terms between the exposure and time!). We also assume that each person was observed for the complete 14 visits. Your complete simulated data will include `simID`, `visit`, `rand`, and all baseline covariates (see Table 3). We then can calculate

the predicted survival `p` at each person-time observation, and survival (`s`) by taking the cumulative product of `p` at each visit. This is equivalent to using the Kaplan-Meier method ($s = s * (1 - p)$) to calculate survival. (*Note:* Cumulative incidence can be calculated here as $ci = 1 - s$.)

(4) **Simulated data: placebo.** Repeat step 3, but assign everyone to placebo. Don't forget to recreate the interaction terms! Note, in Stata we can do Steps 3 and 4 together.

(5) **Calculate average counterfactual survival.** First, concatenate your two simulated datasets into one dataset, only keeping `s`, `visit`, and `rand` in the new dataset called `both`. To calculate average survival at each visit, we have to average over all individuals separately for each new trial arm. The result is a dataframe or dataset called `results`). We then have a dataset that has 30 observations and 3 variables, an estimate of the average `s` for each individual at each `visit` under each hypothetical treatment assignment `rand`. (Note, in Stata we do not create a new dataset but instead expand and add to our current working dataset.)

(6) **Counterfactual survival curves** Plot the survival estimates from Step 5 results.

(7) **Average hazard ratio, risk difference, and risk ratio at 14 weeks.** We can use the counterfactual survival estimates to calculate the average causal effect on the hazard ratio, risk difference, and risk ratio scales. We estimate the hazard ratio at each visit during follow-up as `log(Treated) / log(Placebo)`, and the risk difference and risk ratio using the cumulative incidence (CI = 1-S). Since we have an interaction between randomization and time, our model does not make the proportional hazard assumption, so we have to choose what hazard ratio to estimate. Here, we average the HR over the whole study period [this is just one option].

Original Data

simID	visit	rand	mi_bin	NIHA_b	Y
9	0	1	0	0	0
9	1	1	0	0	0
9	2	1	0	0	0
9	3	1	0	0	0
9	4	1	0	0	0
9	5	1	0	0	0
9	6	1	0	0	0
9	7	1	0	0	1

$\Longrightarrow$

Simulated Data - `rand` = 1

simID	visit	rand	mi_bin	NIHA_b
9	0	1	0	0
9	1	1	0	0
9	2	1	0	0
9	3	1	0	0
9	4	1	0	0
9	5	1	0	0
9	6	1	0	0
9	7	1	0	0
9	8	1	0	0
9	9	1	0	0
9	10	1	0	0
9	11	1	0	0
9	12	1	0	0
9	13	1	0	0
9	14	1	0	0

$+$

Simulated Data - `rand` = 0

simID	visit	rand	mi_bin	NIHA_b
9	0	0	0	0
9	1	0	0	0
9	2	0	0	0
9	3	0	0	0
9	4	0	0	0
9	5	0	0	0
9	6	0	0	0
9	7	0	0	0
9	8	0	0	0
9	9	0	0	0
9	10	0	0	0
9	11	0	0	0
9	12	0	0	0
9	13	0	0	0
9	14	0	0	0

Table 3: Visualization of Simulation Process

Run the code in **Code Section 4**, fill in the table below, and answer the following questions.

`visit` (t)	$\hat{S}^{z=0}(t)$	$\hat{S}^{z=1}(t)$	RD	HR	CIR
0				– – –	– – –
1				– – –	– – –
2				– – –	– – –
3				– – –	– – –
4				– – –	– – –
5				– – –	– – –
6				– – –	– – –
7				– – –	– – –
8				– – –	– – –
9				– – –	– – –
10				– – –	– – –
11				– – –	– – –
12				– – –	– – –
13				– – –	– – –
14				– – –	– – –
15					

Question 1

What is the average risk difference at the end of follow-up? What is the average intention-to-treat hazard ratio at the end of follow-up? What is the average cumulative incidence ratio at the end of follow-up?

Question 2

How does the average hazard ratio compare with the *conditional* unadjusted intention-to-treat hazard ratio? What about the *conditional* covariate adjusted hazard ratio?

Question 3

What is the causal interpretation of this standardized hazard ratio?

Question 4

What assumptions are we making to give this average hazard ratio a causal interpretation?

4 Exercise 3: Estimating per-protocol effects

4.1 Background

The per-protocol effect is the effect of receiving treatment according to the trial protocol (or target trial protocol). Since treatment happens after randomization, the per-protocol effect is not guaranteed to be free from confounding, even in a randomized trial.

In 1980, the Coronary Drug Project team published an analysis comparing the 5-year survival among individuals who did and did not adhere to placebo, and detected a large association between placebo adherence and survival [2]. The results of this study have been used to argue that adherence adjustment (and by extension, per-protocol effect estimation) cannot be done in a randomized trial. However, re-analyses of the Coronary Drug Project accounting for post-randomization confounding using inverse probability weighting showed that this association was spurious [7, 8]. Whenever you have a placebo arm and believe that placebo truly has no effect on the outcome of interest, comparing adherers to non-adherers in the placebo arm can be a good way to assess whether you may have collected enough post-randomization confounders to be able to estimate a per-protocol effect.

Examples of per-protocol effects that can be estimated in survival analyses are:

Additive: Cumulative incidence difference at the end of follow-up

$$\text{RD} = \Pr[Y^{\bar{a}=1} = 1] - \Pr[Y^{\bar{a}=0} = 1]$$

Multiplicative Cumulative incidence ratio at the end of follow-up

$$\text{RR} = \frac{\Pr[Y^{\bar{a}=1} = 1]}{\Pr[Y^{\bar{a}=0} = 1]}$$

Multiplicative Hazard ratio

$$\text{HR} = \frac{\lambda^{\bar{a}=1}}{\lambda^{\bar{a}=0}}$$

In order to estimate the per-protocol effect, we first need to define the protocol of interest. For simplicity, we will specify the **static** protocol: "continuously adhere to assigned treatment or placebo". However, we could also have specified a **dynamic** strategy such as "continuously adhere to assigned treatment or placebo until a pre-specified contraindication develops, after which stop taking assigned treatment or placebo".

Once we have specified our protocol of interest, there are two main approaches we can use to estimate the per-protocol effect – a censoring approach and a dose-response approach. For our protocol we can either:

- Determining whether each individual is adherent or not at baseline and then artificially censoring them if and when they stop adhering.
- Determining whether each individual is adherent or not at every time point and modeling adherence as a continuous variable.

In both approaches, inverse probability weighting is then used to adjust for the potential selection bias induced by the artificial censoring, creating a counterfactual world where everyone either continuously adheres to treatment or placebo. When we use inverse probability weighting, we again need to calculate robust standard errors or use bootstraps to get valid confidence intervals.

In this exercise, we will use the first approach. For an example of the second approach, see [8].

4.2 Data cleaning and exploration

Clear the workspace and read in a new copy of the trial data. We will redo our data cleaning so that we are censoring people once they stop adhering. We do this by creating a new variable that is

0 if an individual's adherence is the same as it was at baseline, or 1 once it has changed. Then we need to make new interaction terms with time, recreate the `maxVisit` variables (so it only includes person-time that is not censored), and subsequently the `deathOverall` variable, and make a new `baseline` subset.

Finally, we will calculate the Kaplan-Meier curve for survival separately in (uncensored) adherers in the treatment and placebo arms.

Run the code in **Code Section 5** and answer the following questions.

Question 1

How many individuals in the placebo arm are adherent at baseline? How many individuals in the treatment arm are adherent at baseline?

Question 2

Comment on the Kaplan-Meier curves you generated. After visually inspecting the data, do the curves look similar?

4.3 Estimating the per-protocol effect

To estimate the per-protocol effect, we need to control for baseline and time-varying (post-randomization) confounding of adherence and mortality. Since many of the confounders may also be affected by prior adherence, we need to use g-methods to ensure that our adjustment doesn't introduce additional bias [4]. In this example, we will use inverse probability of adherence weighting.

To do this, the **Exercise 3 Code** follows these steps:

(1) **Estimate the inverse probability of adherence weights** We will use pooled logistic regression models to estimate the probability of adherence at each visit conditional on only baseline covariates and prior adherence (for the numerator of the weights), and the probability of adherence at each visit conditional on the baseline and post-randomization covariates and prior adherence (for the denominator of the weights). Using the predicted probabilities generated by the weights models, we then calculate the inverse probability of adherence weights. We estimate these weights separately in each trial arm since the reasons for non-adherence may differ between the two arms.

(2) **Estimate the outcome model using a weighted pooled logistic regression model** Estimate the parameters of a pooled logistic regression model. This is exactly the same as the adjusted pooled logistic regression model we ran in Exercise 2, except that we need to restrict our model to run only in the person-time where individuals are continuously adherent and we specify that the model be run in the weighted data (the **pseudo-population**).

(3) **Standardize over baseline covariates** Since we have used stabilized weights, we can also standardize our estimates to generate counterfactual survival curves and obtain marginal interpretations. This standardization is the same as in Exercise 2, except that we now simulate

if everyone in the placebo arm had been continuously adherent and if everyone in the treatment arm had been continuously adherent.

(4) **Counterfactual survival curves** Plot the survival estimates obtained via standardization.

(5) **Average hazard ratio, risk difference, and risk ratio at 14 weeks.** Calculate the average effects of interest from our survival estimates obtained via standardization.

4.3.1 Estimate inverse probability of adherence weights

To estimate the effect of adherence on all-cause mortality in the placebo arm, we first need to create **inverse probability of adherence weights**. These weights will create a pseudo-population where the association between adherence and time-varying confounders is removed, allowing adherence to appear to be randomized with respect to the observed time-varying covariate distribution. There are many methods to calculate weights (unadjusted, stabilized, normalized, truncated). Stabilized weights can be estimated easily whenever we have static sustained treatment strategies or point exposures, but dynamic sustained treatment strategies often require unstabilized weights.

Unstabilized weights are calculated based on the predicted probability of an individual having the adherence (or exposure) history they actually received, conditional on their baseline and time-varying covariates: ω_{it} shows the formula for calculating an individual's unstabilized weight up to time t.

$$\omega_{it} = \prod_{j=0}^{t} \frac{1}{f_D(A_{ij} \mid L_0, L_t, \bar{A}_{ij-1}, Z = z)}$$

Stabilized weights are similar to unstabilized weights but have as a numerator the predicted probability of an individual having the adherence (or exposure) history they actually received, conditional on their baseline covariates only: $s\omega_{it}$ shows the formula for calculating an individual's stabilized weight up to time t.

$$s\omega_{it} = \prod_{j=0}^{t} \frac{f_N(A_{ij} \mid L_0, \bar{A}_{ij-1}, Z = z)}{f_D(A_{ij} \mid L_0, L_t, \bar{A}_{ij-1}, Z = z)}$$

In this example we will use **truncated** stabilized weights – that is, stabilized weights which are truncated to prevent extreme values.

To calculate stabilized weights, we follow these steps in each trial arm:

- Fit the model for the numerator: estimate the probability of adherence at time t conditional on baseline covariates only

$$\begin{aligned} logit(\Pr[A_t = 1 \mid A_0, L_0, t > 0, Z = z]) &= \gamma_{0,t} + \gamma_1 A_0 + \gamma_2 L_0 \\ &= \gamma_0^* + \gamma_1 A_0 + \gamma_2 L_0 + \gamma_3 t + \gamma_4 t^2 \end{aligned}$$

- Fit the model for the denominator: estimate the probability of adherence at time t conditional on baseline *and time-varying* covariates

$$\begin{aligned} logit(\Pr[A_t = 1 \mid A_0, L_0, L_t, t > 0, Z = z]) &= \delta_{0,t} + \delta_1 A_0 + \delta_2 L_0 + \delta_3 L_t \\ &= \delta_0^* + \delta_1 A_0 + \delta_2 L_0 + \delta_3 L_t + \delta_4 t + \delta_5 t^2 \end{aligned}$$

- From each model, predict the probability of adherence at time t in the data.

– Calculate the appropriate contribution to the weight from each model fit.

$$f_N(A_{it} \mid L_0, Z = z) = A_{it}\Pr(A_{it} = 1|Z = z) + (1 - A_{it})\Pr(A_{it} = 0|Z = z)$$
$$f_D(A_{it} \mid L_0, L_t, Z = z) = A_{it}\Pr(A_{it} = 1|Z = z) + (1 - A_{it})\Pr(A_{it} = 0|Z = z)$$

– Take the product of the ratio of the numerator and denominator contributions up to time t^* to calculate the stabilized weights at time t^*. Truncate if needed (usually at the 99^{th} percentile).

You may notice that we have fit the weights models only in the post-baseline visits. This is because we want to include prior adherence as a predictor in the model and at baseline we don't have prior adherence. The way we have structured our weights, we are not controlling for baseline confounding at this stage. However, we can control for baseline confounding directly in the outcome model by including our baseline covariates there – in fact, we would have to do this even if we had modeled weights for basleine because we stabilized our weights by the baseline covariates.

Also, note that in other settings, we may want to include prior adherence (or exposure) history in the weights models. For simplicity, and because we have little adherence switching, baseline adherence is a reasonable control for prior adherence history in our models..

Run the code in **Code Section 6**, fill in the table, and answer the following questions.

	Unstabilized	Stabilized	Truncated Stabilized (99^{th} percentile)
Mean (SD)			
Range			
Median (IQR)			
99th Percentile			

Question 1

What was the mean of your stabilized weights after truncation? What did you expect it to be? Did stabilizing the weights (compared to unstabilized) change the mean of the weights? How did truncating change the mean and range of the weights?

4.3.2 Estimating the conditional hazard ratio

Now, we can use the stabilized inverse probability of adherence weights to estimate the hazard ratio for overall mortality for our `per-protocol` effect, where our protocols are 1) continually adhere to the protocol ("take at least 80% of assigned placebo pills"), versus 2) continually adhere to the treatment ("take at least 80% of assigned treatment pills"). We will censor individuals when they deviate from their protocol (we can do this based on `maxVisit` or `adhr_t` $\neq$ `1`).

We will use a pooled logistic regression model to estimate this per-protocol effect. Since we are using a censoring based approach (1 from above), our outcome model will be created in the person-time where adherence status is always high from baseline (`t` $\leq$ `maxVisit` or `adhr_t = 1`). We then use randomization status (`adhr_b`) as our exposure and the baseline covariates included above (since we stabilized over them) as covariates in the model. We don't need to include adherence in the model because everyone in our dataset after censoring has the same value of adherence at all time points.

$$logit[\Pr(Y_{t+1} = 1 \mid Z, \bar{A}_t = 1, L_0, \bar{Y}_t = 0)] = \beta_0 + \beta_1 Z + \beta_2 L_0 = \beta_0^* + \beta_1 Z + \beta_2 L_0 + \beta_3 t + \beta_4 t^2$$

Run the code in **Code Section 7** and answer the following questions. Here, we will only consider a stabilized truncated weighting procedure (`stabw_t`). If you would like to see the instability from unstabilized or stabilized weights, replace the weights argument with `unstabw` or `stabw`.

`adhr_b`	Stabilized (Truncated)
Coef.	
SE	
Hazard Ratio	

Question 1

What is the estimated conditional hazard ratio using stabilized (truncated) weights? How do you interpret this hazard ratio?

Question 2

What assumptions does your interpretation rely upon?

Question 3

How does this compare to the intention-to-treat estimate? Is this what you expected?

4.3.3 Estimating the average survival curves

Next, we can estimate the average survival if everyone adhered to each treatment. To do this, we could use the same weighted pooled logistic regression model as above, but to allow for greater flexibility in estimating the baseline hazard, we choose to refit this model including a product term between randomization (`rand`) and time (`visit`) to our outcome model.

To standardize our estimate so we can have marginal estimates, we will use the same approach as we used when we were estimating the standardized ITT (see Code Section 4).

Note: You may be wondering where the time-varying covariates and adherence have gone. We only use the weights to get the correct coefficients from the outcome regression model, we do not need weights to do any predicting. For prediction, we only need the variables on the right-hand side of the model statement, which in our case are all time-fixed (except for time and its spline term, which we can create ourselves in the simulated data), and don't include adherence. Therefore, we can take the same strategy as in Code Section 4 to create our simulated dataset.

Run the code in **Code Section 8**, fill in the table, and answer the following questions. Here, $S^0(t)$ refers to non-adherers, and $S^1(t)$ refers to adherers.

`visit` (t)	$S^0(t)$	$S^1(t)$	RD	HR	CIR
0				– – –	– – –
1				– – –	– – –
2				– – –	– – –
3				– – –	– – –
4				– – –	– – –
5				– – –	– – –
6				– – –	– – –
7				– – –	– – –
8				– – –	– – –
9				– – –	– – –
10				– – –	– – –
11				– – –	– – –
12				– – –	– – –
13				– – –	– – –
14				– – –	– – –
15					

Question 1

What is the average hazard ratio for the placebo arm? What is the risk difference at the end

of follow-up? The cumulative incidence ratio at the end of follow-up?

Question 2

How do these compare to the intention-to-treat effect estimates? Is this what you expected?

Question 3

How do you interpret the standardized survival curves?

Question 4

What assumptions does your interpretation rely upon?

References

[1] Coronary Drug Project Research Group. The coronary drug project: design, methods, and baseline results. *Circulation*, 47(3suppl):I1–50, 1973.

[2] Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *New England Journal of Medicine*, 303(18):1038–1041, 1980.

[3] Coronary Drug Project Research Group et al. Clofibrate and niacin in coronary heart disease. *J Am Med Assoc*, 231:360–381, 1975.

[4] Miguel A Hernán and James M Robins. *Causal Inference*. Chapman & Hall/CRC, forthcoming, 2019.

[5] Miguel A Hernán, James M Robins, et al. Per-protocol analyses of pragmatic trials. *N Engl J Med*, 377(14):1391–1398, 2017.

[6] Eleanor J Murray, Ellen C Caniglia, Sonja A Swanson, Sonia Hernández-Díaz, and Miguel A Hernán. Patients and investigators prefer measures of absolute risk in subgroups for pragmatic randomized trials. *Journal of clinical epidemiology*, 103:10–21, 2018.

[7] Eleanor J Murray and Miguel A Hernán. Adherence adjustment in the coronary drug project: a call for better per-protocol effect estimates in randomized trials. *Clinical Trials*, 13(4):372–378, 2016.

[8] Eleanor J Murray and Miguel A Hernán. Improved adherence adjustment in the coronary drug project. *Trials*, 19(1):158, 2018.

[9] Wikipedia contributors. Clofibrate. — Wikipedia, the free encyclopedia, 2004. [Online; accessed 22-Feb-2019].