Causal Inference with R: Module 2 Adjusting for Confounding via Standardization and Difference-in-Differences

Babette Brumback

Professor Emerita
University of Florida, Biostatistics Department

Assumptions

- These methods require the assumption that H is a sufficient confounder for estimating the effect of T on Y; that is,

$$\{Y(t)\}_{t \in \mathcal{T}} \amalg T | H, \tag{1}$$

  where $\mathcal{T}$ is the set of possible values for T, as determined using the Backdoor Theorem explained previously.
- We also require the positivity assumption that

$$1 > P(T = t|H) > 0$$

  for all possible values t of T.
- Finally, we require the consistency assumption.

Standardization via Outcome Modeling

- Standardization via outcome modeling gives us a way to estimate $E(Y(t))$ assuming (1), positivity, and consistency. We can write

$$E(Y(t)) = E_H E(Y(t)|H) =$$
$$E_H E(Y(t)|T = t, H) = E_H E(Y|T = t, H), \quad (2)$$

  where the first equality follows from the double expectation theorem, the second follows from assumption (1), which implies mean independence of $Y(t)$ and T given H, and the third follows from the consistency assumption, which allows us to replace $Y(t)$ with Y when we condition on $T = t$.
- $E(Y|T = t, H)$ is known as the outcome model.
- Note that

$$E_H E(Y|T = t, H) \neq E_{H|T=t} E(Y|T = t, H) = E(Y|T = t),$$

  where $E_{H|T=t} E(Y|T = t) = E(Y|T = t)$ by the double expectation theorem.

Standardization via Outcome Modeling

- Mistakenly equating $E_H E(Y|T=t|H)$ with $E_{H|T=t}E(Y|T=t,H)$ would imply that $E(Y(t)) = E(Y|T=t)$, which would only be true if $Y(t) \amalg T$.
- However, we have assumed confounding by H; that is, $Y(t) \amalg T|H$.
- Equation (2) is the basis for the outcome-modeling approach to standardization.
- If H is continuous or high-dimensional, we cannot express $E(Y|T=t,H)$ nonparametrically, but rather must use a parametric outcome model.
- With a binary dataset, we can write

$$E_H E(Y|T=t,H) = E(Y|T=t,H=0)P(H=0) + E(Y|T=t,H=1)P(H=1).$$

We can thus estimate $E(Y(t))$ using nonparametric estimates of the components on the right-hand side of the equation.

Standardization via Outcome Modeling

- We consider some examples. For the mortality data, recalling that $T = 1$ indicates the US and $H = 1$ indicates the 65+ age group, we have that

$$\begin{aligned}\hat{E}(Y|T=0,H=0) &= 0.002254\\ \hat{E}(Y|T=0,H=1) &= 0.05652\\ \hat{E}(Y|T=1,H=0) &= 0.002679\\ \hat{E}(Y|T=1,H=1) &= 0.04460\end{aligned} \tag{3}$$

and

$$\hat{P}(H=0) = \frac{282{,}305{,}227+1{,}297{,}258{,}493}{282{,}305{,}227+1{,}297{,}258{,}493+48{,}262{,}955+133{,}015{,}479} = 0.897$$
$$\hat{P}(H=1) = 1 - \hat{P}(H=0) = 1 - 0.897 = 0.103$$

- Therefore,

$$\hat{E}(Y(0)) = 0.002254 * 0.897 + 0.05652 * 0.103 = 0.0078434$$
$$\hat{E}(Y(1)) = 0.002679 * 0.897 + 0.04460 * 0.103 = 0.0069969$$

Standardization via Outcome Modeling

- Thus we see that when we combine the age-specific mortality rates of China with the age distribution of the two countries combined, we compute an overall mortality rate of around 7.8 per 1000, whereas when we do the same for the US, we compute about 7.0 per thousand.
- The standardized risk difference is thus -0.8 per 1000, and the standardized relative risk is 0.897.
- This is a reversal from the unadjusted comparison of the two countries, which was in the opposite direction, at 8.8 per 1000 versus 7.3 per 1000, with a risk difference of 1.2 per 1000 and a relative risk of 1.206.
- Due to the large sample sizes per country, the reversal is highly statistically significant.

Standardization via Outcome Modeling

- Next, we use R to compute standardized estimates for the What-If? study. We assumed that

$$\{Y(0), Y(1)\} \amalg A|H.$$

- We can use the nonparametric linear model

$$E(Y|A, H) = \beta_1 + \beta_2 A + \beta_3 H + \beta_4 A * H$$

to estimate

$$E(Y(0)) = E_H E(Y|A = 0, H) = \beta_1 + \beta_3 E(H)$$

and

$$E(Y(1)) = E_H E(Y|A = 1, H) = \beta_1 + \beta_2 + \beta_3 E(H) + \beta_4 E(H).$$

Standardization via Outcome Modeling

Table 1: Standardized Estimates for the What-If? Study

Measure	Estimate	95% CI
$\hat{E}(Y(0))$	0.375	(0.263, 0.487)
$\hat{E}(Y(1))$	0.289	(0.206, 0.372)
$\hat{RD}$	-0.086	(-0.212, 0.039)
$\hat{RR}$	0.770	(0.528, 1.12)

Standardization via Outcome Modeling

- For our final example, we analyze data from the Double What-If? Study. Using the causal DAG and the `R` code for `doublewhatifsim.r` that was used to simulate the data, we can find that

$$\{VL_1(0), VL_1(1)\} \amalg A | AD_0,$$

but that

$$\{VL_1(0), VL_1(1)\} \not\amalg A | VL_0.$$

We can also find that

$$E(VL_1(0)) = 0.655$$
$$E(VL_1(1)) = 0.295$$

- For estimation, we can reuse the R code used for the What-If? Study, labeling $Y = VL_1$ and $H = AD_0$.

Standardization via Outcome Modeling

Table 2: Standardized Estimates for the Double What-If? Study with $H = AD_0$

Measure	Estimate	95% CI
$\hat{E}(VL_1(0))$	0.669	(0.636, 0.702)
$\hat{E}(VL_1(1))$	0.335	(0.272, 0.397)
$\hat{RD}$	-0.334	(-0.403, -0.265)
$\hat{RR}$	0.500	(0.413, 0.606)

- We observe that the confidence interval for $E(VL_1(0))$, which is (0.636, 0.702), includes the true value 0.655; the confidence interval for $E(VL_1(1))$, which is (0.272,0.397), includes the true value 0.295.

Standardization via Outcome Modeling

- As expected, the method appears to work. However, without knowledge of the foregoing proof that the method does indeed work, we might wonder if we were just lucky with this particular sample, which we obtained with `set.seed(444)`.
- More empirical evidence that the method works can be obtained by repeatedly running `doublewhatifsim.r` without the `set.seed(444)` command in order to generate 1000 confidence intervals; we can then ascertain whether the percentage of confidence intervals covering the true values is 95%.
- We implemented this to obtain more evidence that the method is valid.
- However, one simulation study is not a general proof. Fortunately, we have already proved that the method works.

Standardization via Outcome Modeling

- For comparison, we repeat the standardization with $H = VL_0$.

Table 3: Standardized Estimates for the Double What-If? Study with $H = VL_0$

Measure	Estimate	95% CI
$\hat{E}(VL_1(0))$	0.696	(0.662, 0.729)
$\hat{E}(VL_1(1))$	0.245	(0.192, 0.299)
$\hat{RD}$	-0.450	(-0.512, -0.388)
$\hat{RR}$	0.353	(0.283, 0.441)

Standardization via Outcome Modeling

- As VL_0 is not a sufficient confounder, we do not expect these results to be correct.
- Indeed, the confidence interval for $E(VL_1(0))$, which is (0.662,0.729), does not include the true $E(VL_1(0))$, which is 0.655.
- However, the confidence interval for $E(VL_1(1))$, (0.272,0.397), does include the true $E(VL_1(1))$, which is 0.295.
- Therefore, the results are biased, but not terribly so.
- Furthermore, due to sampling variability, it is possible that the method is valid but that our confidence interval belongs to the 5% of the 95% confidence intervals that do not contain the true value.
- We conducted a simulation study and found that the percentage of confidence intervals for $E(VL_1(0))$ covering 0.655 is 63.3%, while the percentage of confidence intervals for $E(VL_1(1))$ covering 0.295 is 13.7%.
- Because these percentages are so far away from 95%, the simulation study indicates substantial bias.
- It is important to keep in mind that this is only one simulation study, and that other simulation studies with different data generating mechanisms might reveal more or less bias.
- However, just one simulation study demonstrating bias is enough to prove that standardization with an insufficient confounder, such as VL_0, is not generally valid.

Average Effect of Treatment on the Treated

- Previously, we introduced a special kind of conditional causal effect, called the *average effect of treatment on the treated* (ATT), i.e. $E(Y(1)|A=1)$ versus $E(Y(0)|A=1)$.
- We can estimate $E(Y(1)|A=1)$ easily, because it equals $E(Y|A=1)$ by consistency.
- We can use an outcome-modeling approach to standardization to estimate $E(Y(0)|A=1)$, assuming (1), with A in place of T, and consistency:

$$E(Y(0)|A=1) = E_{H|A=1}E(Y(0)|A=1,H) = E_{H|A=1}E(Y(0)|A=0,H) = E_{H|A=1}E(Y|A=0,H), \quad (4)$$

  where the first equality follows from the double expectation theorem, the second from (1), and the third from consistency.
- For binary H, we can estimate the outcome model non-parametrically, as before, but for continuous or high-dimensional H, we can enlist a parametric model for $E(Y|A=0,H)$.
- We can use a similar argument to estimate $E(Y(1)|A=0)$; this is left as an exercise.

Average Effect of Treatment on the Treated

- For the mortality data, letting $A = T$, the ATT compares mortality in the US versus China using the age distribution in the US.
- Recalling that $P(H = 1|A = 1) = 0.146$, so that $P(H = 0|A = 1) = 0.854$, and using the mortality rates at (3), we compute $E(Y(0)|A = 1)$ from (4) as

$$E(Y|A = 0, H = 0)P(H = 0|A = 1) +$$
$$E(Y|A = 0, H = 1)P(H = 1|A = 1) =$$
$$0.002254 * 0.854 + 0.05652 * 0.146 = 0.0102.$$

- Thus, had the age distribution been the same as in the US, the mortality rate in China would have been 10.2 per thousand instead of 7.3 per thousand.
- We can compute $E(Y(1)|A = 1)$ directly from the US data as $E(Y|A = 1) = 0.0088$, or 8.8 per thousand.
- The ATT is thus 8.8 per thousand versus 10.2 per thousand.
- For the What-If? study, we use `R` to estimate the ATT, with a slight modification to our previous code.

Average Effect of Treatment on the Treated

Table 4: ATT for the What-If? Study

Measure	Estimate	95% CI
$\hat{E}(Y(0)\|A=1)$	0.361	(0.247, 0.476)
$\hat{E}(Y(1)\|A=1)$	0.276	(0.192, 0.360)
$\hat{RD}$	-0.085	(-0.207, 0.037)
$\hat{RR}$	0.765	(0.520, 1.12)

- In this example, we find that the results reported in Table 4 are very similar to the overall average effect of treatment, $E(Y(1)) - E(Y(0))$, computed previously, although the results for both are variable due to the relatively small sample size.
- This suggests that any effect modifiers are balanced across the two treatment groups.

Average Effect of Treatment on the Treated

- For comparison with difference-in-differences estimation presented next, we estimate the ATT for the Double What-If? Study, first with $H = AD_0$ and second with $H = VL_0$.
- The analysis with $H = AD_0$ is correct, whereas the one with $H = VL_0$ is not.
- Computing the true values of $E(VL_1(1)|A = 1)$, $E(VL_1(0)|A = 1)$, and the ATT using the true data generating mechanisms in `doublewhatifsim.r` is difficult but not impossible.
- We can derive that

$$E(VL_1(0)|A = 1) = 0.559.$$

  and that

$$E(VL_1(1)|A = 1) = 0.199.$$

- Thus, we have $E(VL_1(0)|A = 1) = 0.559$, $E(VL_1(1)|A = 1) = 0.199$, the true RD equal to $E(VL_1(1)|A = 1) - E(VL_1(0)|A = 1) = -0.36$, and the true RR equal to $E(VL_1(1)|A = 1)/E(VL_1(0)|A = 1) = 0.356$.

Average Effect of Treatment on the Treated

Table 5: Standardized ATT Estimates for the Double What-If? Study with $H = AD_0$

Measure	Estimate	95% CI
$\hat{E}(VL_1(0)\|A = 1)$	0.574	(0.526, 0.622)
$\hat{E}(VL_1(1)\|A = 1)$	0.231	(0.179, 0.283)
$\hat{RD}$	-0.344	(-0.404, -0.283)
$\hat{RR}$	0.402	(0.322, 0.501)

Average Effect of Treatment on the Treated

- We observe that the confidence interval for $E(VL_1(0)|A = 1)$, which is (0.526, 0.622), includes the true value 0.559, and the one for $E(VL_1(1)|A = 1)$, which is (0.179, 0.283), includes the true value 0.199.
- The true RD, -0.36, is in its respective confidence interval (-0.404,-0.283), and the true RR, 0.356, also falls in its respective confidence interval (0.322, 0.501).
- As expected, the method appears to work.
- We conducted a simulation study to find that the confidence intervals cover $E(Y(0)|A = 1)$ 93.7% of the time, $E(Y(0)|A = 1)$ 94.7% of the time, the true RD 95.3% of the time, and the true RR 96% of the time.
- This supports the validity of the method.
- The next slide presents results for $H = VL_0$.

Average Effect of Treatment on the Treated

Table 6: Standardized ATT Estimates for the Double What-If? Study with $H = VL_0$

Measure	Estimate	95% CI
$\hat{E}(VL_1(0)\|A=1)$	0.682	(0.647, 0.718)
$\hat{E}(VL_1(1)\|A=1)$	0.231	(0.177, 0.284)
$\hat{RD}$	-0.452	(-0.514, -0.389)
$\hat{RR}$	0.338	(0.267, 0.428)

- As VL_0 is not a sufficient confounder, we do not expect these results to be correct.
- Indeed, the only confidence intervals that include the true values are for $E(VL_1(1)|A=1)$ and RR. For the other two measures, the results are biased, but not terribly so. We should not expect the confidence interval for $E(VL_1(1)|A=1)$ and the RR to cover the true value in general.
- Once again, we conducted a simulation study and found that the confidence intervals cover $E(VL_1(0)|A=1)$ 0% of the time, $E(VL_1(1)|A=1)$ 96.0% of the time, the true RD 7.7% of the time, and the true RR 75.5% of the time. Therefore, for the data generating model of `doublewhatifsim.r`, use of VL_0 rather than AD_0 as a sufficient confounder is invalid for three of the four measures.

Standardization with a Parametric Outcome Model

- When H necessitates a parametric outcome model, validity of the standardization depends upon correct model specification. In practice, our model will never be exactly right, but hopefully it is close enough that our results are not too far off.
- We present two examples.
- First, we analyze the What-If? data in `whatif2dat`, which includes some continuous covariates.

Standardization with a Parametric Outcome Model

```
> head(whatif2dat)
  vl0 vlcont0 artad0 vl4 vlcont4 artad4 audit0 T A lvlcont0 lvlcont4
1   0      20      1   1     420      1      1 0 1   2.9957   6.0403
2   0      20      0   0      20      1      1 1 1   2.9957   2.9957
3   1   61420      1   0      20      1      1 1 0  11.0255   2.9957
4   1     600      0   0      20      1      1 1 1   6.3969   2.9957
5   1   75510      0   1  184420      0      1 0 1  11.2320  12.1250
6   0      20      1   0      30      0      1 1 0   2.9957   3.4012
```

Standardization with a Parametric Outcome Model

- The variable `lvlcont0` represents log of viral load (copies/ml) at baseline, a continuous variable that we will use as H in the analysis of the effect of reduced drinking `A` on unsuppressed viral load at four months `vl4`.
- The following `R` code computes the estimates that follow.

Standardization with a Parametric Outcome Model

```
> bootstand.r
function ()
{
stand.out<-boot(data=whatif2dat,statistic=standout.r,R=1000)
stand.est<-summary(stand.out)$original
stand.SE<-summary(stand.out)$bootSE
stand.lci<-stand.est-1.96*stand.SE
stand.uci<-stand.est+1.96*stand.SE
list(stand.est=stand.est,stand.SE=stand.SE,stand.lci=stand.lci,
stand.uci=stand.uci)
}
> standout.r
function(data=whatif2dat,ids=c(1:nrow(whatif2dat)))
{
dat<-data[ids,]
lmod<-glm(vl4~A+lvlcont0,family=binomial,data=dat)
dat0<-dat1<-dat
dat0$A<-0
dat1$A<-1
EYhat0<-predict(lmod,newdata=dat0,type="response")
EYhat1<-predict(lmod,newdata=dat1,type="response")
EY0<-mean(EYhat0)
EY1<-mean(EYhat1)
rd<-EY1-EY0
logrr<-log(EY1/EY0)
c(EY0,EY1,rd,logrr)
}
```

Standardization with a Parametric Outcome Model

Table 7: Outcome-model Standardization for the What-If? Study with $H =$ `lvlcont0`

Measure	Estimate	95% CI
$\hat{E}(Y(0))$	0.360	(0.249, 0.472)
$\hat{E}(Y(1))$	0.300	(0.216, 0.384)
$\hat{RD}$	-0.061	(-0.188, 0.067)
$\hat{RR}$	0.831	(0.564, 1.23)

- We observe that the results of Table 1 based on the nonparametric outcome model with H equal to unsuppressed viral load at baseline are quite similar to those of Table 7 based on the parametric outcome model with H equal to the log of viral load at baseline.

Standardization via Exposure Modeling

- Standardization via exposure modeling gives us a second way to estimate $E(Y(t))$ assuming (1), positivity, and consistency.
- We present the method for binary T. To estimate $E(Y(t))$ for non-binary T, one can first recode the data so that $T = 1$ when it previously equaled t, and $T = 0$ when it previously equaled any value other than t.
- Then one can use the method for estimating $E(Y(1))$ with the recoded data.
- The exposure model is $E(T|H) = P(T = 1|H)$, so named because sometimes T indicates a potentially harmful exposure, rather than a treatment.
- The exposure model is also known as the *propensity score*, denoted by $e(H)$, as it is a function of H. It is called the propensity score because it measures the propensity for treatment given observed levels of the confounders H.
- The most common parametric model for it is the logistic model

$$E(T|H) = \text{expit}(\alpha_0 + H_1\alpha_1 + \cdots + H_q\alpha_q), \tag{5}$$

where q is the number of components of H.

Standardization via Exposure Modeling

- One can prove the following relations, which then can be used to estimate $E(Y(0))$ and $E(Y(1))$.

$$E(Y(1)) = E\left(\frac{TY}{e(H)}\right) \tag{6}$$

and

$$E(Y(0)) = E\left(\frac{(1-T)Y}{1-e(H)}\right). \tag{7}$$

Standardization via Exposure Modeling

- For better understanding, we will estimate $E(Y(1))$ and $E(Y(0))$ using the mortality data and empirical versions of (6) and (7), and compare with the outcome-modeling standardization results.

Standardization via Exposure Modeling

```
> mk.mortdat
function(){
mortdat<-NULL
mortdat$H<-c(0,0,0,0,1,1,1,1)
mortdat$T<-c(0,0,1,1,0,0,1,1)
mortdat$Y<-c(0,1,0,1,0,1,0,1)
mortdat$n<-c((1297258493-2923480),2923480,
(282305227-756340),756340,(133015479-7517520),7517520,(48262955-2152660),
2152660)
mortdat$p<-mortdat$n/sum(mortdat$n)
eH0<-sum(mortdat$n[3:4])/sum(mortdat$n[1:4])
eH1<-sum(mortdat$n[7:8])/sum(mortdat$n[5:8])
mortdat$eH<-eH0*(1-mortdat$H)+ eH1*mortdat$H
mortdat$s1<-mortdat$T*mortdat$Y/mortdat$eH
mortdat$s0<-(1-mortdat$T)*mortdat$Y/(1-mortdat$eH)
EY1<-sum(mortdat$s1*mortdat$p)
EY0<-sum(mortdat$s0*mortdat$p)
mortdat<-data.frame(mortdat)
list(EY1=EY1,EY0=EY0,mortdat=mortdat)
}
```

Standardization via Exposure Modeling

```
> mortdat.out<-mk.mortdat()
> mortdat.out
$EY1
[1] 0.0069952
$EY0
[1] 0.0078399
$mortdat
  H T Y          n          p      eH     s1     s0
1 0 0 0 1294335013 0.73506589 0.17872 0.0000 0.0000
2 0 0 1    2923480 0.00166027 0.17872 0.0000 1.2176
3 0 1 0  281548887 0.15989445 0.17872 0.0000 0.0000
4 0 1 1     756340 0.00042953 0.17872 5.5952 0.0000
5 1 0 0  125497959 0.07127156 0.26624 0.0000 0.0000
6 1 0 1    7517520 0.00426928 0.26624 0.0000 1.3628
7 1 1 0   46110295 0.02618650 0.26624 0.0000 0.0000
8 1 1 1    2152660 0.00122252 0.26624 3.7561 0.0000
```

Standardization via Exposure Modeling

- In the function `mk.mortdat`, which makes the `mortdat` dataset and calculates the estimated potential outcomes, `eH` is $e(H)$, `s1` is $TY/e(H)$, and `s0` is $(1-T)Y/(1-e(H))$.
- To compute the estimated potential outcomes, we need to sum the `s1` or `s0` summands weighted by the probability of the row, `p`.
- We see that $\hat{E}(Y(1)) = 0.0069952$ and $\hat{E}(Y(0)) = 0.0078399$, which would be identical to the estimates obtained earlier using the outcome-modeling approach, except for propagation of round-off error.

Average Effect of Treatment on the Treated

- We can also use the exposure-modeling approach to estimate the ATT. We introduce $e_0 = P(T = 1)$, to go along with $e(H) = P(T = 1|H)$.
- We showed in equation (4), letting $T = A$, that

$$E(Y(0)|T = 1) = E_{H|T=1}E(Y|T = 0, H). \tag{8}$$

- One can prove that $E(Y(0)|T = 1)$ is also a function of the exposure model. Specifically,

$$E(Y(0)|T = 1) = E\left(\frac{Y(1 - T)e(H)}{(1 - e(H))e_0}\right). \tag{9}$$

Average Effect of Treatment on the Treated

- We compute $E(Y(0)|T=1)$ as follows

```
> attsem.r
function(mortdat=mortdat.out$mortdat)
{
e0<-sum(mortdat$T*mortdat$p)
s<-mortdat$Y*(1-mortdat$T)*mortdat$eH/(e0*(1-mortdat$eH))
EY0T1<-sum(s*mortdat$p)
EY0T1
}
> attsem.r()
[1] 0.010176
```

- We find that $\hat{E}(Y(0)|T=1) = 0.0102$, or 10.2 per thousand, identical to the estimate computed using the outcome-modeling approach. $E(Y(1)|T=1)$ can be estimated via $E(Y|T=1)$, as before, at 8.8 per thousand.

Standardization with a Parametric Exposure Model

- For parametric exposure models such as (5), equation (6) allows us to estimate $E(Y(1))$ by first estimating α with $\hat{\alpha}$ using an estimating equation for logistic models, second computing $\hat{e}(H)$ with $\hat{\alpha}$, and third computing

$$\hat{E}(Y(1)) = \frac{1}{n}\Sigma_i \frac{T_i Y_i}{\hat{e}(H_i)}. \tag{10}$$

- Similarly, we can estimate $E(Y(0))$ as

$$\hat{E}(Y(0)) = \frac{1}{n}\Sigma_i \frac{(1 - T_i) Y_i}{1 - \hat{e}(H_i)}. \tag{11}$$

Standardization with a Parametric Exposure Model

- Another good estimator of $E(Y(1))$ is given by

$$\hat{E}(Y(1)) = \frac{\Sigma_i \frac{T_i Y_i}{\hat{e}(H_i)}}{\Sigma_i \frac{T_i}{\hat{e}(H_i)}}.$$

- This takes the form of a random *weighted average*,

$$\hat{E}(Y(1)) = \Sigma_i W_i Y_i, \quad \text{where}$$
$$W_i = \frac{\frac{T_i}{\hat{e}(H_i)}}{\Sigma_i \frac{T_i}{\hat{e}(H_i)}}, \tag{12}$$

  where $W_i \in [0, 1]$ and $\Sigma_i W_i = 1$.
- Note that $(1/n)\Sigma_i Y_i$ is a weighted average with $W_i = 1/n$. Similarly,

$$\frac{\Sigma_i a_i Y_i}{\Sigma_i a_i}$$

  is a weighted average for any nonnegative a_i.

Standardization with a Parametric Exposure Model

- We also have that

$$\hat{E}(Y(0)) = \Sigma_i W_i Y_i, \quad \text{where}$$
$$W_i = \frac{\frac{1-T_i}{1-\hat{e}(H_i)}}{\Sigma_i \frac{1-T_i}{1-\hat{e}(H_i)}}. \tag{13}$$

- The estimators of (12) and (13) are useful for estimation using the `weighted.mean` function or the `glm` or `geeglm` functions with the `weights` options.
- Note that the denominators of the weights are constant for a given sample, and our use of the three `R` functions is invariant to multiplication of the weights by such a constant. Therefore, in our data examples, we use only the numerators for the weights.

Standardization with a Parametric Exposure Model

- First, we apply exposure-model standardization to the What-If? Study with $H =$ `lvlcont0`, for comparison with the outcome-modeling results reported in Table 7.
- We use the `standexp.r` function shown below, noting that the use of the `weights` option in the `glm` function produces the correct estimator of `beta`, but that the standard errors are incorrect for our usage.
- Therefore, we turn to the bootstrap, which is also convenient for computing confidence intervals for functions of the parameters, including the relative risk.
- Our `bootstand.r` function is the same except for the modification to `statistic=standexp.r`.

Standardization with a Parametric Exposure Model

```
> standexp.r
function(data,ids)
{
dat<-data[ids,]
e<-fitted(glm(A~lvlcont0,family=binomial,data=dat))
dat$W<-(1/e)*dat$A + (1/(1-e))*(1-dat$A)
beta<-glm(vl4~A,data=dat,weights=W)$coef
EY0<-beta[1]
EY1<-beta[1]+beta[2]
rd<-EY1-EY0
rr<-log(EY1/EY0)
c(EY0,EY1,rd,rr)
}
```

Standardization with a Parametric Exposure Model

Table 8: Exposure-model Standardization for the What-If? Study with $H =$ `lvlcont0`

Measure	Estimate	95% CI
$\hat{E}(Y(0))$	0.360	(0.249, 0.471)
$\hat{E}(Y(1))$	0.300	(0.220, 0.380)
$\hat{RD}$	-0.060	(-0.188, 0.069)
$\hat{RR}$	0.834	(0.565, 1.23)

Standardization with a Parametric Exposure Model

- If interest is focused primarily on the risk difference, one could alternatively program its estimator and standard error using `exp.r`.

```
> exp.r
function(dat=whatif2dat)
{
library(geepack)
e<-fitted(glm(A~lvlcont0,family=binomial,data=dat))
dat$W<-(1/e)*dat$A + (1/(1-e))*(1-dat$A)
dat$ids<-c(1:nrow(dat))
summary(geeglm(vl4~A,data=dat,id=ids,weights=W))
}
> exp.r()
Call:
geeglm(formula = vl4 ~ A, data = dat, weights = W, id = ids)

 Coefficients:
            Estimate Std.err  Wald Pr(>|W|)
(Intercept)   0.3600  0.0613 34.47  4.3e-09
A            -0.0598  0.0767  0.61     0.44
```

Standardization with a Parametric Exposure Model

- We see that the coefficient of `A`, -0.0598, is identical to our estimate of the RD in Table 8.
- It has been shown that for large samples, the standard error, 0.0767, of the RD estimated with `exp.r` is necessarily larger than our bootstrap estimate of the standard error estimated with `bootstand.r` and `standexp.r`.
- As the latter standard error is 0.0655, the rule holds for this example.
- Because the standard error estimated with `exp.r` is too large, if one were to find statistical signficance, typically P $<$ 0.05, then one would not need to do any further analyses, because the bootstrap P-value would be even smaller.
- However, if P $>$ 0.05, as it does in this example with $P = 0.44$, one would typically wish to check statistical signficance with the bootstrap.
- Our bootstrap confidence interval includes 0, which means that for this example, both methods produce statistically insignificant results.

Doubly Robust Standardization

- Validity of standardization with a parametric outcome or exposure model requires correctness of the chosen model.
- The two methods can yield quite different results.
- Doubly robust standardization avoids this problem by using an estimator that relies on both models and that is valid if at least one of the models is correct.
- The method is based on the following relations, that are true assuming (1), consistency, positivity, and that either the exposure model $e(H)$ or the outcome model $E(Y|H,T)$ is correctly specified.

$$E(Y(1)) = E\left(\frac{TY}{e(H)} - \frac{T-e(H)}{e(H)}E(Y|H,T=1)\right), \tag{14}$$

and

$$E(Y(0)) = E\left(\frac{(1-T)Y}{1-e(H)} + \frac{T-e(H)}{1-e(H)}E(Y|H,T=0)\right). \tag{15}$$

Doubly Robust Standardization

- Using (14) and (15), we can estimate $E(Y(1))$ and $E(Y(0))$ with

$$\hat{E}(Y(1)) = (1/n)\left\{\Sigma_i \left(\frac{T_i}{\hat{e}(H_i)} Y_i - \frac{T_i - \hat{e}(H_i)}{\hat{e}(H_i)} \hat{E}(Y_i | H_i, T_i = 1)\right)\right\} \qquad (16)$$

and an analogous expression for $\hat{E}(Y(0))$.

Doubly Robust Standardization

- To illustrate, we apply the method to the What-If? Study data assuming `lvlcont0` is a sufficient confounder for the effect of `A` on `vl4`.
- We use a misspecified outcome model including only `A` and the intercept, and we use an exposure model including the intercept and `lvlcont0`.
- The results of standardization using only the exposure model were shown previously in Table 8, and the results of standardization using only the misspecified outcome model are shown below in Table 9.

Doubly Robust Standardization

Table 9: Outcome-model Standardization for the What-If? Study with the Misspecified Outcome Model

Measure	Estimate	95% CI
$\hat{E}(Y(0))$	0.400	(0.277, 0.523)
$\hat{E}(Y(1))$	0.276	(0.190, 0.362)
$\hat{RD}$	-0.124	(-0.273, 0.025)
$\hat{RR}$	0.690	(0.442, 1.07)

- Focusing on the risk difference, the estimate and 95% confidence interval are -0.060 (-0.188,0.069) using the exposure model as compared to -0.124 (-0.273, 0.025) using the misspecified outcome model.
- Although neither method leads to statistical significance, because both confidence intervals include zero, the estimate using the misspecified outcome model suggests a larger effect that is close to statistical significance.

Doubly Robust Standardization

```
> badstanddr.r
function(data,ids)
{
dat<-data[ids,]
e<-fitted(glm(A~lvlcont0,family=binomial,data=dat))
lmod<-glm(vl4~A,family=binomial,data=dat)
dat0<-dat1<-dat
dat0$A<-0
dat1$A<-1
EYhat0<-predict(lmod,newdata=dat0,type="response")
EYhat1<-predict(lmod,newdata=dat1,type="response")
EY0<-mean(dat$vl4*(1-dat$A)/(1-e) + EYhat0*(e-dat$A)/(1-e))
EY1<-mean(dat$vl4*(dat$A/e) - EYhat1*(dat$A-e)/e)
rd<-EY1-EY0
rr<-log(EY1/EY0)
c(EY0,EY1,rd,rr)
}
```

Doubly Robust Standardization

Table 10: Doubly Robust Standardization for the What-If? Study Combining the Misspecified Outcome Model of Table 9 and the Exposure Model of Table 8

Measure	Estimate	95% CI
$\hat{E}(Y(0))$	0.362	(0.253, 0.471)
$\hat{E}(Y(1))$	0.300	(0.216, 0.385)
$\hat{RD}$	-0.062	(-0.183, 0.060)
$\hat{RR}$	0.830	(0.571, 1.20)

- The estimate and 95% confidence interval for the risk difference are -0.062 (-0.183, 0.060), quite close to those of the exposure-modeling approach, which are themselves almost identical to those estimated with the outcome model using both `A` and `lvlcont0`, previously presented in Table 7.
- This should come as no surprise, as we saw previously that the conditional expectation of the outcome given `A` and `lvlcont0` depends on `lvlcont0`.

Difference-in-Differences Estimation: Introduction

- On March 11, 2020, the World Health Organization (WHO) declared COVID-19 a pandemic. The public health crisis affected countries across the globe, with several, including the US, shutting down large segments of their economies in order to stem the spread of the coronavirus.
- In the US, estimated monthly employment rates for men and women from January to August 2020 are reported on the next slide.
- These estimates were produced by the US Bureau of Labor Statistics using data from the Current Population Survey, which is a monthly survey of households conducted by the Bureau of Census.
- The estimates are seasonally adjusted, meaning that increases or decreases from month to month are due to factors other than seasonal variation for a typical year.
- These data document, in a very obvious way, that shutting down the economy subsequent to the pandemic declaration triggered a large rise in unemployment.
- We see a relatively small rise from February to March, followed by a large rise in April and beyond.
- In this example, no one would question that the increase was caused by the nation's response to the pandemic.
- In the face of cause and effect this obvious, sophisticated statistical methods are not really needed. Simple subtraction, i.e. 16.2% - 4.4% = 11.8% for women or 13.5% - 4.4% = 9.1% for men, is enough.

Table 11: US Monthly Percent Unemployment from January to August 2020 of the Civilian Noninstitutional Population 16 Years and Over, Estimated by the Bureau of Labor Statistics Using Current Population Survey Data

Month	Percent of Men	Percent of Women
January	3.6	3.5
February	3.6	3.4
March	4.4	4.4
April	13.5	16.2
May	12.2	14.5
June	10.6	11.7
July	9.8	10.6
August	8.3	8.6

Introduction

- In this example, we were able to rule out seasonal variation as a major confounder for two reasons.
- First, the estimates were reported as seasonally adjusted, meaning that this confounder has already been removed, perhaps via standardization.
- Second, the magnitude of the differences is too great to be due to seasonal fluctuations alone.
- For many other investigations involving pre- versus post-exposure differences, cause and effect is not as obvious.

Introduction

- For example, Molyneux et al. (2019) estimated the effect of negative interest rate policy (NIRP) on bank margins and profits.
- Responding to the global financial crisis of 2007-2008, the central banks of many countries implemented NIRP in order to provide economic stimulus to weak economies.
- Former President Trump repeatedly tweeted about the benefits of negative interest rates. For example, on September 3, 2019, @realDonaldTrump tweeted "Germany, and so many other countries, have negative interest rates, 'they get paid for loaning money,' and our Federal Reserve fails to act! Remember, these are also our weak currency competitors!"
- Many economists are not as sanguine.
- Molyneux et al. (2019) analyzed a dataset comprising 7,359 banks from 33 OECD member countries over 2012-2016 to assess the impact of NIRP on net interest margins (NIMs).

Introduction

- NIM measures the net amount a bank earns on loans and other interest-earning assets relative to the amount of those loans and other assets.
- For example, supposing the bank's interest earning assets equal one million dollars in a year, the bank earned \$50,000 in interest at 5% and paid \$20,000 in expenses to their lenders, then the NIM would be (\$50,000-\$20,000)/\$1,000,000=3%.
- Table 2 of Molyneux et al. (2019) reports on yearly NIMs of banks in countries initiating NIRP both pre-NIRP and post-NIRP.
- There are 8916 bank-years (one bank-year represents one NIM from one bank from one year) pre-NIRP with an average NIM of 2.06% and a standard deviation of 0.95% (hence a standard error of $0.95/\sqrt{8916} = 0.0100\%$) and 8040 bank-years post-NIRP with an average NIM of 1.92% and a standard deviation of 0.78% (hence a standard error of $0.78/\sqrt{8040} = 0.0087\%$).
- Ignoring the temporal correlation between NIMs of a single bank from year to year, we can assess whether the difference in average NIMs, that is 1.92%-2.06% = -0.14%, is statistically significant using a z-test, computing

$$z = \frac{-0.14}{\sqrt{0.0100^2 + 0.0087^2}} = -10.562,$$

  which indicates that the difference is highly statistically significant.
- However, it may be due to factors other than initiation of NIRP. Perhaps temporal changes in other variables led to the difference over time.

Introduction

- The difference-in-differences solution incorporates a control group, resulting in one of the most popular tools for applied research in economics to evaluate the effects of public interventions and other treatments of interest on relevant outcome variables.
- We compare the change over time in the exposed group to the change over time in the unexposed group.
- Molyneux et al. (2019) also present statistics on banks in countries that did not initiate NIRP, over a matched time period.
- There are 4686 control bank-years pre-NIRP with an average NIM of 2.92% and a standard deviation of 1.71% (hence a standard error of $1.71/\sqrt{4686} = 0.0250\%$) and 4331 control bank-years post-NIRP with an average NIM of 2.93% and a standard deviation of 1.65% (hence a standard error of $1.65/\sqrt{4331} = 0.0251\%$).
- We use the following `R` code to determine whether the difference in differences, that is, $(1.92 - 2.06) - (2.93 - 2.92) = -0.15\%$, is statistically significant.

```
> analyze.r
function ()
{
did<-(2.06-1.92)-(2.92-2.93)
se<-sqrt((.95^2)/8916 + (.78^2)/8040 + (1.71^2)/4686 + (1.65^2)/4331)
list(did=did,se=se,z=did/se)
}
> analyze.r()
$did
[1] 0.15
$se
[1] 0.037809
$z
[1] 3.9673
>2*pnorm(-3.9673)
[1] 7.2691e-05
```

Introduction

- The z-statistic is 3.97, corresponding to a P-value less than 0.0001, indicating that the difference in differences is indeed statistically significant, which suggests that initiation of NIRP negatively affects banks.
- Next, we provide the foundation for this approach to adjusting for confounding.
- We note that Molyneux et al. (2019) also applied a more complicated difference-in-differences approach that accounted for the temporal correlation of NIMs from a single bank and also adjusted for other factors, but their results were qualitatively the same as ours.

Difference-in-Differences (DiD) Estimators

- Let Y_t, $t = 0, 1$, denote the pre- and post-exposure measures.
- Let A indicate the exposure.
- Let $Y_1(0)$ and $Y_1(1)$ denote the potential post-exposure outcomes to $A = 0$ and $A = 1$, respectively.

DiD Estimator with a Linear Model

- The method relies on consistency as well as assumption A1:

$$\text{A1:}\quad E(Y_1(0)|A=1) - E(Y_1(0)|A=0) = E(Y_0|A=1) - E(Y_0|A=0), \quad (17)$$

  also called additive equi-confounding.
- The target of estimation is the linear ATT, presented previously as a risk difference, but also valid for non-binary Y_t:

$$\text{Linear ATT:}\quad E(Y_1(1) - Y_1(0)|A=1). \quad (18)$$

DiD Estimator with a Linear Model

- The DiD estimator derives from the relation

$$E(Y_1(1) - Y_1(0)|A = 1) = E(Y_1|A = 1) - E(Y_1|A = 0) - (E(Y_0|A = 1) - E(Y_0|A = 0)), \quad (19)$$

  which connects the estimand framed in terms of potential outcomes to an estimand relying only on observed data.

- To prove the relation, note that

$$E(Y_1|A = 1) - E(Y_1|A = 0) = E(Y_1(1)|A = 1) - E(Y_1(0)|A = 0),$$

  by consistency, and furthermore that

$$E(Y_0|A = 1) - E(Y_0|A = 0) = E(Y_1(0)|A = 1) - E(Y_1(0)|A = 0),$$

  by assumption A1.

- Taking the difference of these two differences proves the relation.

DiD Estimator with a Linear Model

- We can therefore estimate the linear ATT at (18) via the difference in differences of averages,

$$\hat{E}(Y_1|A=1) - \hat{E}(Y_1|A=0) - (\hat{E}(Y_0|A=1) - \hat{E}(Y_0|A=0)),$$

  which equals

$$\hat{E}(Y_1 - Y_0|A=1) - \hat{E}(Y_1 - Y_0|A=0). \tag{20}$$

- This is the estimator we used for the NIRP example.

DiD Estimator with a Linear Model

- We can also compute the DiD estimator via the linear model

$$E(Y_t|A) = \alpha_0 + \alpha_1 t + \alpha_2 A + \beta A * t, \tag{21}$$

where

$$\begin{aligned}\beta &= E(Y_1|A=1) - E(Y_1|A=0) - (E(Y_0|A=1) - E(Y_0|A=0)) \\ &= (\alpha_0 + \alpha_1 + \alpha_2 + \beta) - (\alpha_0 + \alpha_1) \\ &\quad - ((\alpha_0 + \alpha_2) - \alpha_0)\,.\end{aligned}$$

- Thus β is the linear ATT, and we can use linear regression to estimate it.
- We will see examples in `R` in a bit.

DiD Estimator with a Loglinear Model

- The method relies on consistency as well as assumption A2:

$$\text{A2:} \quad E(Y_1(0)|A=1)/E(Y_1(0)|A=0) = E(Y_0|A=1)/E(Y_0|A=0), \tag{22}$$

  which one might also call multiplicative equi-confounding or additive equi-confounding on the log scale.
- The target of estimation is the loglinear ATT:

$$E(Y_1(1)|A=1)/E(Y_1(0)|A=1). \tag{23}$$

DiD Estimator with a Loglinear Model

- The DiD estimator derives from the relation

$$\log E(Y_1(1)|A=1) - \log E(Y_1(0)|A=1) =$$
$$\log E(Y_1|A=1) - \log E(Y_1|A=0) - (\log E(Y_0|A=1) - \log E(Y_0|A=0)),$$

  which connects the estimand framed in terms of potential outcomes to an estimand relying only on observed data.
- The proof is similar to that for relation (19), relying on consistency and assumption A2.

DiD Estimator with a Loglinear Model

- We can therefore estimate the log of the loglinear ATT at (23) via the difference in differences of log averages, analogous to (20), and then exponentiate.
- We can also compute the DiD estimator via the loglinear model

$$\log E(Y_t|A) = \alpha_0 + \alpha_1 t + \alpha_2 A + \beta A * t, \tag{24}$$

where

$$\begin{aligned}\beta &= \log E(Y_1|A=1) - \log E(Y_1|A=0) - (\log E(Y_0|A=1) - \log E(Y_0|A=0)) \\ &= (\alpha_0 + \alpha_1 + \alpha_2 + \beta) - (\alpha_0 + \alpha_1) \\ &\quad - ((\alpha_0 + \alpha_2) - \alpha_0)\,.\end{aligned}$$

- Thus β is the log of the loglinear ATT, and we can use loglinear regression to estimate it. We then exponentiate the results.

DiD Estimator with a Logistic Model

- The method relies on consistency as well as assumption A3:

$$\text{A3:} \qquad \begin{aligned} &\text{logit}E(Y_1(0)|A=1) - \text{logit}E(Y_1(0)|A=0) = \\ &\qquad \text{logit}E(Y_0|A=1) - \text{logit}E(Y_0|A=0), \end{aligned}$$

  which one might also call additive equi-confounding on the logit scale.
- The target of estimation is the logistic ATT:

$$\text{logit}E(Y_1(1)|A=1) - \text{logit}E(Y_1(0)|A=1). \tag{25}$$

DiD Estimator with a Logistic Model

- The DiD estimator derives from the relation

$$\text{logit}E(Y_1(1)|A=1) - \text{logit}E(Y_1(0)|A=1) =$$
$$\text{logit}E(Y_1|A=1) - \text{logit}E(Y_1|A=0) - (\text{logit}E(Y_0|A=1) - \text{logit}E(Y_0|A=0)),$$

  which connects the estimand framed in terms of potential outcomes to an estimand relying only on observed data.
- The proof is similar to that for relation (19), relying on consistency and assumption A3.

DiD Estimator with a Logistic Model

- We can therefore estimate the logistic ATT at (25) via the difference in differences of logit averages, analogous to (20).
- For binary Y_t, exponentiating yields an odds ratio.
- We can also compute the DiD estimator via the logistic model using the regression

$$\text{logit}E(Y_t|A) = \alpha_0 + \alpha_1 t + \alpha_2 A + \beta A * t, \tag{26}$$

  where β is the logistic ATT, and we can use logistic regression to estimate it.
- Again, for binary Y_t, exponentiating yields an odds ratio.

Comparison with Standardization

- The DiD approach to adjusting for confounding relies on consistency plus one of assumptions A1, A2, or A3, which involve the confounder Y_0.
- We saw previously that we can also use standardization to estimate the linear, loglinear, or logistic ATT.
- However, standardization relies on consistency plus the sufficient confounder assumption.
- Interestingly, and as we will see below in the case of the Double What-If? Study, when Y_0 is not a sufficient confounder, it can still happen that one of assumptions A1, A2, or A3 is true.
- This implies that a DiD analysis may be valid even when standardization using only Y_0 may not be.
- Conversely, Y_0 may be a sufficient confounder without A1, A2, or A3 holding.
- In that case, standardization using only Y_0 would be valid whereas the DiD analyses would not be.
- It is worth noting that when one of A1, A2, or A3 holds, the other two will typically not hold.
- Therefore, the validity of the DiD approach depends on correctly choosing which one might plausibly hold, often an impossible task.
- However, in our analysis of the Double What-If? Study, we observe that the confidence intervals for all three estimators include the true value for their respective estimands.

Comparison with Standardization

- In our exposition of the three DiD approaches, we assumed Y_0 was a pre-exposure version of the post-exposure Y_1. More generally, we could let Y_0 represent another variable, say H, that could reasonably be presumed to satisfy one of the assumptions A1, A2, or A3, and the corresponding DiD estimator would be plausibly valid. For example, supposing a set of baseline covariates X predictive of the post-treatment outcome Y_1, H could represent the prognostic score, $E(Y_1|X)$.
- For binary datasets, the estimand corresponding to the DiD estimator with a linear model can be expressed as

$$\begin{aligned} &E(Y_1|A=1) - (E(Y_1|A=0,Y_0=1) - E(Y_1|A=0,Y_0=0))\, E(Y_0|A=0) \\ &\quad -E(Y_1|A=0,Y_0=0) - (E(Y_0|A=1) - E(Y_0|A=0))\,, \end{aligned} \tag{27}$$

by substituting

$$\begin{aligned} E(Y_1|A=0) = E(Y_1|A=0,Y_0=0)(1-E(Y_0|A=0)) \\ +E(Y_1|A=0,Y_0=1)E(Y_0|A=0) \end{aligned}$$

into the relation at (19).

Comparison with Standardization

- On the other hand, the estimand corresponding to the standardized ATT can be expressed as

$$E(Y_1|A=1) - (E(Y_1|A=0,Y_0=1) - E(Y_1|A=0,Y_0=0))\,E(Y_0|A=1) \\ -E(Y_1|A=0,Y_0=0). \tag{28}$$

- Subtracting (28) from (27) yields

$$(E(Y_1|A=0,Y_0=1) - E(Y_1|A=0,Y_0=0) - 1)\,(E(Y_0|A=1) - E(Y_0|A=0))\,. \tag{29}$$

- Therefore, the two estimands will differ unless either (a) $Y_1 = Y_0$ for everyone with $A=0$, or (b) $Y_0 \amalg A$.
- In the first case, the ATT equals $E(Y_1 - Y_0|A=1)$, and DiD estimation is unnecessary. In the second case, the ATT equals $E(Y_1|A=1) - E(Y_1|A=0)$, and standardization is unnecessary.
- Because the estimands will typically differ, we recommend trying both approaches whenever possible, as it is unlikely that scientific considerations will be able to distinguish which approach has more validity in a given context.

Comparison with Standardization

- Next we consider an example.
- Recall that the Double What-If? Study was simulated according to `doublewhatifsim.r`. The pre- and post-exposure measures are VL_0 and VL_1, which take the place of Y_0 and Y_1 above.
- Previously, we derived that $E(VL_1(0)|A = 1) = 0.559$ and $E(VL_1(1)|A = 1) = 0.199$.

Comparison with Standardization

- From the code, we can derive that

$$E(VL_1 - VL_0|A = 1) - E(VL_1 - VL_0|A = 0) = -0.36.$$

- Therefore, the expected value of the DiD estimator with a linear model equals -0.36.
- From the lecture on standardizaiton, we have that that the linear ATT also equals $E(VL_1(1)|A = 1) - E(VL_1(0)|A = 1) = -0.36$, which means that the DiD estimator with a linear model is valid in this example.
- It is not likely that the DiD estimators with the loglinear model and the logistic model are valid, but we present these for the sake of comparison.

R Code

- We next present the `R` code we used to compute the DiD estimators, which relied on the regression model formulations of (21), (24), and (26).
- First, we needed to transform the `doublewhatifdat` dataset from *short form* into *long form*.
- Long form contains one row per person per time period, whereas short form contains one row per person, which includes measures at both time periods.

R Code

```
> mklongdwi.r
function(dat=doublewhatifdat)
{
longdat<-NULL
for (i in 1:nrow(dat))
{
VL<-dat[i,"VL0"]
A<-dat[i,"A"]
time<-0
longdat<-rbind(longdat,c(VL,A,time))
VL<-dat[i,"VL1"]
A<-dat[i,"A"]
time<-1
longdat<-rbind(longdat,c(VL,A,time))
}
dimnames(longdat)[[2]]<-c("VL","A","time")
data.frame(longdat)
}
> longdwidat<-mklongdwi.r()
> head(longdwidat)
  VL A time
1  1 0    0
2  1 0    1
3  1 0    0
4  1 0    1
5  1 0    0
6  1 0    1
```

R Code

- Next we compute β for each of the models using `did.r`, and we estimate confidence intervals using `bootdid.r`.
- We need to bootstrap `doublewhatifdat` rather than `longdwidat` so that we take a bootstrap sample of the participants rather than of the time points.
- For that reason, we call `mklongdwi.r` from within `did.r`.

R Code

```
> did.r
function(data,ids)
{
dat<-data[ids,]
dat<-mklongdwi.r(dat)
beta<-lm(VL~A+time+A*time,data=dat)$coef
rd<-beta[4]
beta<-glm(VL~A+time+A*time,family=poisson,data=dat)$coef
logrr<-beta[4]
beta<-glm(VL~A+time+A*time,family=binomial,data=dat)$coef
logor<-beta[4]
c(rd,logrr,logor)
}
> bootdid.r
function ()
{
stand.out<-boot(data=doublewhatifdat,statistic=did.r,R=1000)
stand.est<-summary(stand.out)$original
stand.SE<-summary(stand.out)$bootSE
stand.lci<-stand.est-1.96*stand.SE
stand.uci<-stand.est+1.96*stand.SE
list(stand.est=stand.est,stand.SE=stand.SE,stand.lci=stand.lci,stand.uci=stand.uci)
}
```

Results

Table 12: Difference-in-Differences Estimation of the ATT for the Double What-If? Study

Measure	Truth	Estimate	95% CI
RD	-0.360	-0.355	(-0.439, -0.272)
RR	0.356	0.400	(0.313, 0.512)
OR	0.196	0.206	(0.139, 0.307)

Connection to Lord's Paradox

- Lord's Pardox pertains to unrestricted pre-treatment Y_0 and post-treatment Y_1 variables and the so-called ANCOVA model

$$E(Y_1|Y_0, A) = \beta Y_0 + \alpha_1 A + \alpha_0,$$

where furthermore

$$E(Y_0|A = 1) = (\alpha_1 + \alpha_0)/(1 - \beta)$$

and

$$E(Y_0|A = 0) = \alpha_0/(1 - \beta),$$

which render the difference-in-differences (DiD) estimand equal to zero, i.e.

$$E(Y_1 - Y_0|A = 1) = E(Y_1 - Y_0|A = 0) = 0$$

.

- We observe that it can also pertain to binary Y_1 and Y_0 as long as β, α_1, and α_0 are such that $E(Y_1|Y_0, A)$ and $E(Y_0|A)$ belong to [0,1].

Connection to Lord's Paradox

- The DiD assumption can be expressed as

$$E(Y_1(0)|A=1) = E(Y_0|A=1) - E(Y_0|A=0) + E(Y_1|A=0).$$

For unrestricted Y_0 and Y_1, this assumption is compatible with any distribution of the observed data (Y_0, A, Y_1). For binary Y_0 and Y_1, it is only compatible when the right-hand side belongs to [0,1].
Under the DiD assumption and the ANCOVA model, the average effect of treatment on the treated (ATT) estimand is the DiD estimand, which equals zero.

- Under the standardization assumption

$$(Y_1(0), Y_1(1)) \amalg A|Y_0$$

and the ANCOVA model, the ATT estimand equals

$$E_{Y_0}E(Y_1|Y_0, A=1) - E_{Y_0}E(Y_1|Y_0, A=0),$$

which equals

$$\beta E(Y_1(0)) + \alpha_1 + \alpha_0 - \beta(E(Y_1(0)) + \alpha_0 = \alpha_1.$$

- Therefore, unless $\alpha_1 = 0$, the DiD and standardization estimators will conflict with one another. Moreover, either estimator could be valid, unless the outcome is restricted and the expression for $E(Y_1(0)|A=1)$ is out of bounds.

In Closing

- In closing, DiD estimation is an attractive method to adjust for confounding in settings involving comparisons of pre- and post-exposure measures Y_0 and Y_1 or, more generally, of pre-exposure prognostic scores H and post-exposure measures Y_1.
- The general DiD assumption can be expressed as $E(Y_1(0)|A=1) = E(H|A=1) - E(H|A=0) + E(Y_1|A=1)$. For binary datasets, the DiD assumption could only be valid when the right-hand side belongs to [0,1]. For unrestricted outcomes, it could always be valid. When the DiD assumption is valid, one can estimate $E(Y_1(0)|A=1)$ with an estimate of the right-hand side and combine it with an estimate of $E(Y_1|A=1)$ to estimate any of the linear, loglinear, and logistic ATT.
- Standardization could also be valid in any setting.
- Lord's Paradox is that DiD estimation and ANCOVA estimation (which, in his example, is equivalent to estimation of the ATT using standardization) can lead to conflicting results. We show that there is no good resolution to the paradox regarding estimation of the causal average effect of treatment on the treated; assuming the ANCOVA model holds, DiD estimation is valid when the DiD assumption holds, whereas ANCOVA estimation is valid when the stratified randomized trial assumption $(Y_1(0), Y_1(1)) \amalg A|Y_0$ holds.
- When both approaches could be valid, we recommend trying each of them to make sure that results are qualitatively similar.
- If they are not similar, then caution is warranted in interpreting results, and the study may be deemed inconclusive.

Extra Material: Doubly Robust Standardization

- It is natural to question when one should consider using standardization with an outcome model versus with an exposure model versus doubly robust standardization.
- When the first two methods give discrepant answers, doubly robust standardization might help choose between them.
- However, if we choose doubly robust standardization in the first place, we might pay a price in terms of variability of the estimate.
- One might think that standardization with the exposure model would be preferable when the outcome indicates a rare condition.
- To see this, first suppose the condition is not rare.
- We might have 3000 individuals and 50%, or 1500, with the condition.
- Using the rule of thumb for logistic regression presented in Chapter 2, we should be able to include 150 covariates in the outcome model.
- Now suppose the condition is rare, and 1%, or 30, have it. Our rule of thumb now suggests we can only include 3 covariates in the outcome model.
- If our sufficient confounder, H, is high-dimensional, what do we do?
- Suppose the exposure, T, is divided more evenly: that is, we have 600 with $T = 1$. This would suggest we can include 60 covariates in the exposure model. Would it not be preferable to use only the exposure model for standardization?

Doubly Robust Standardization

- To investigate, we turned to the simulation study `simdr.r` with `ss` indicating the number of columns of H.
- The columns of H were independent indicator variables each with probability 0.05.
- We simulated T as indicator variables with probabilities that varied as a linear function of H, such that approximately 600 individuals had $T = 1$.
- We simulated Y_i as a function of T_i and $\Sigma_{k=1}^{\texttt{ss}} H_{ik}$, such that approximately 35 individuals had $Y = 1$.
- The mean of $\Sigma_{k=1}^{\texttt{ss}} H_{ik}$ was fixed at one, but when `ss` was set to 100, it ranged from 0.00 to 2.80.
- $P(T = 1|H)$ ranged from 0.041 to 0.468.
- $E(Y|T, H)$ ranged from 0.000 to 0.036.
- The range of $\hat{e}(H_i)$ fitted with a correctly specified linear exposure model was -0.037 to 0.575.
- Because the propensity score should not be negative, we also tried a logistic exposure model for exposure model standardization, even though it was not correctly specified.

Doubly Robust Standardization

- We let `ss`, which directly relates to the number of covariates in the model, equal 40 and 100.
- We would have expected exposure-model standardization to be best with 40, and we would have expected all methods to break down with 100.
- However, this did not happen.
- To try to force the outcome-modeling and doubly robust standardization to break down, we overspecified the outcome model, using

  ```
  Y~T*H
  ```

  to include all two-way interactions between T and the columns of H.

Doubly Robust Standardization

```
> simdr.r
function(ss=100)
{
H<-matrix(0,3000,ss)
probH<-rep(0.05,3000)
for (i in 1:ss)
{
H[,i]<-rbinom(n=3000,size=1,prob=probH)
}
sumH<-apply(H,1,sum)*(20/ss)
#return(range(sumH))
probT<-.13*sumH + .05*rnorm(n=3000,mean=1,sd=.1)
#return(range(probT))
T<-rbinom(n=3000,size=1,prob=probT)
#return(sum(T))
probY<-.01*T + .01*sumH
#return(range(probY))
Y<-rbinom(n=3000,size=1,prob=probY)
#return(sum(Y))
e<-fitted(lm(T~H))
e2<-predict(glm(T~H,family=binomial),type="response")
#return(range(e))
w0<-(1-T)/(1-e)
w1<-T/e
w02<-(1-T)/(1-e2)
w12<-T/e2
```

Doubly Robust Standardization

```
dat0<-dat1<-dat<-data.frame(cbind(Y,T))
dat0$T<-0
dat1$T<-1
out<-lm(Y~T*H)
EY0out<-mean(predict(out,newdata=dat0))
EY1out<-mean(predict(out,newdata=dat1))
EY0exp<-weighted.mean(Y,w=w0)
EY1exp<-weighted.mean(Y,w=w1)
EY0exp2<-weighted.mean(Y,w=w02)
EY1exp2<-weighted.mean(Y,w=w12)
EY0dr<-mean(w0*Y + predict(out,newdata=dat0)*(T-e)/(1-e))
EY1dr<-mean(w1*Y - predict(out,newdata=dat1)*(T-e)/e)
EYT0=mean(Y*(1-T))
EYT1=mean(Y*T)
list(EY0exp=EY0exp,EY1exp=EY1exp,EY0exp2=EY0exp2,
EY1exp2=EY1exp2,EYT0=EYT0,EYT1=EYT1,
EY0out=EY0out,EY1out=EY1out,EY0dr=EY0dr,EY1dr=EY1dr)
}
```

Doubly Robust Standardization

- The results of the simulation study that collected the results of running `simdr.r` 1000 times, in order to obtain the sampling distribution of the estimators, are presented in Tables 13 and 14.
- The unadjusted estimator `EYT0` estimates $E(Y|T=0)$, and `EYT1` estimates $E(Y|T=1)$.
- The estimators `EY0exp` and `EY1exp` use exposure model standardization with the correctly specified linear exposure model, whereas `EY0exp2` and `EY1exp2` use exposure model standardization with the incorrect logistic exposure model.
- The estimators `EY0out` and `EY1out` use outcome model standardization with the overspecified outcome model.
- The doubly robust estimators `EY0dr` and `EY1dr` use the overspecified outcome model and the correctly specified linear exposure model.
- The mean and standard deviation columns show the means and standard deviations of the sampling distributions of the estimators.
- The P-value column tests whether our sample of 1000 estimators comes from a sampling distribution with a true mean equal to $E(Y(0)) = 0.01$ or $E(Y(1)) = 0.02$, whichever is relevant.

Doubly Robust Standardization

Table 13: Sampling Distribution of Estimators from Simulation Study Investigating Small-Sample Robustness: True $E(Y(0)) = 0.01$, True $E(Y(1)) = 0.02$, with 40 Columns of H

Estimator	Description	Mean	Standard Deviation	P-Value
`EYT0`	Unadjusted	0.0076	0.0015	0.00
`EYT1`	Unadjusted	0.0042	0.0012	0.00
`EY0exp`	Linear Exposure Model	0.0100	0.0021	0.92
`EY1exp`	Linear Exposure Model	0.0195	0.0127	0.19
`EY0exp2`	Logistic Exposure Model	0.0101	0.0021	0.42
`EY1exp2`	Logistic Exposure Model	0.0204	0.0064	0.07
`EY0out`	Overspecified Outcome Model	0.0100	0.0021	0.79
`EY1out`	Overspecified Outcome Model	0.0200	0.0066	0.84
`EY0dr`	Doubly Robust	0.0100	0.0021	0.82
`EY1dr`	Douby Robust	0.0197	0.0106	0.37

Doubly Robust Standardization

Table 14: Estimated Sampling Distribution of Estimators from Simulation Study Investigating Small-Sample Robustness: True $E(Y(0)) = 0.01$, True $E(Y(1)) = 0.02$, with 100 Columns of H

Estimator	Description	Mean	Standard Deviation	P-Value
`EYT0`	Unadjusted	0.0079	0.0016	0.00
`EYT1`	Unadjusted	0.0038	0.0012	0.00
`EY0exp`	Linear Exposure Model	0.0100	0.0020	0.61
`EY1exp`	Linear Exposure Model	0.0196	0.0562	0.81
`EY0exp2`	Logistic Exposure Model	0.0100	0.0020	0.73
`EY1exp2`	Logistic Exposure Model	0.0200	0.0068	0.96
`EY0out`	Overspecified Outcome Model	0.0100	0.0020	0.74
`EY1out`	Overspecified Outcome Model	0.0200	0.0069	0.74
`EY0dr`	Doubly Robust	0.0100	0.0020	0.72
`EY1dr`	Douby Robust	0.029	0.1891	0.14

Doubly Robust Standardization

- The only estimators with P-values < 0.05 are the unadjusted estimators `EYT0` and `EYT1`.
- Indeed, we observe from the means of their sampling distribution, 0.0076 and 0.0042, that the unadjusted estimators are biased for the true values, 0.01 and 0.02.
- This confirms that standardization is necessary.
- The standard deviations of the sampling distribution indicate how variable the estimators are; smaller standard deviations are desirable, provided bias is negligible.
- We see that the standard deviation of the sampling distribution is larger for the correctly specified linear exposure model than it is for the incorrectly specified logistic model, which happens to exhibit negligible bias.
- Surprisingly, the standard deviations for the overspecified outcome model are about the same as for the logistic exposure model, despite that the outcome is rare.

Doubly Robust Standardization

- Finally, we see large standard deviations for the doubly robust estimators of $E(Y(1))$, particularly for 100 columns of H.
- The doubly robust estimator appears to be the first to break down as we move from 40 to 100 confounders, with the correctly specified linear exposure model second.
- In conclusion, all of the methods were robust with 40 confounders, and even with 100 confounders, results were not horrible. In particular, we question the logic that might lead to a choice of exposure-model standardization over outcome-model standardization with the outcome indicating a rare condition. Furthermore, we caution against using the doubly robust approach with high-dimensional confounders.