HW4: Noncompliance

STAT 597: Causal Inference

Kari Lock Morgan

Due: 4/1/20 at 1:25pm

SOLUTIONS

Type of Face Mask

One of the (many) problems facing our country currently in light of the current pandemic is the lack of available face masks. Face masks are imperative for health care workers treating those with COVID-19, yet are in short supply - it is estimated that the country currently needs 25 times more masks than it currently has stockpiled. If masks were readily available, they would also be useful for slowing spread in the general public, but mask wearing is currently discouraged for the general public to save the masks for the health care workers who are putting their lives on the line to help others, and so desperately need this critical form of protection.

We are currently facing a severe shortage of *medical* face masks, but might placing a cloth mask over your face be just as effective? Today we'll look at a study[1] comparing the efficacy of cloth masks to medical masks in health care workers. They also looked at a control group who followed their usual practice, which may have included mask wearing of either type, but we'll focus here on the medical masks versus the cloth masks.

The study took place in hospitals in Vietnam, and the units were health care workers working full-time in selected high risk wards. Importantly, the study was a randomized trial. It was actually a cluster randomized trial, randomizing at the level of the hospital ward rather than the individual, but we do not have data on the clustering so ignore this here. Participants were instructed to use their randomized type of mask on every shift for 4 consecutive weeks. However, not everyone complied with this assignment.

The study recorded three different primary outcomes, all binary, whether the health-care worker developed:

- Clinical respiratory illness (CRI)
- influenza-like illness (ILI)
- laboratory-confirmed respiratory virus infection

The coronavirus would likely fall in all three of these categories, but the paper is explicit that the laboratory-confirmed outcome included detection for severe acute respiratory syndrome (SARS) associated coronavirus (SARS-CoV), and coronaviruses 229E, NL63, OC43 and HKU1 (the novel coronavirus we're currently facing did not yet exist at the time of the study).

You can find the paper here.

Report the following for the ILI outcome (but feel free to look at the others as well, if you're interested!):

1) [2] Do you expect relatively good covariate balance here? Why or why not?

Yes, because this is a randomized experiment.

[1] MacIntyre CR, Seale H, Dung TC, et alA cluster randomised trial of cloth masks compared with medical masks in healthcare workersBMJ Open 2015;5:e006577. doi: 10.1136/bmjopen-2014-006577

2) [2] Table 1 in the paper looks at covariate balance. Does balance look relatively good, or do you see anything that might be concerning?

The only covariate that stood out to me was the average number of hand-washings per day - it was quite a bit higher in the medical mask group, which could be concerning because hand washing is certainly helpful for protecting against infectious diseases.

3) [2] Table 2 shows the outcome data for an intent-to-treat analysis. Calculate $\hat{ITT}_Y$. Interpret this estimate in context.

```
#itt.y = 28/580 - 43/569 #CRI
itt.y = 1/580 - 13/569
itt.y
```

```
## [1] -0.02112296
```

4) [2] The study defined compliance as wearing their assigned mask for at least 70% of working hours during the course of the study. Compliance rates for each group are given in Figure 3. Note that this is a different type of noncompliance then we've dealt with in class, because noncompliers do not receive the opposite treatment, but rather move to their own treatment group - not consistently wearing any mask at all. (I thought the relevance to current events outweighed this slightly different framework, and it's good to be able to adapt what we've learned to new situations!). Define **Z** and **W** in this context.

$Z_i = 1$ if randomized to medical mask, $Z_i = 0$ if randomized to cloth mask.

$W_i = 1$ if wore medical mask at least 70% of the time, $W_i = 0$ if wore cloth mask at least 70% of the time, and $W_i = N$ if did not wear the assigned mask at least 70% of the time.

5) [3] Define the possible compliance statuses for this framework (hint: there are four without additional assumptions, but they will differ from what we discussed in class).

- Complier under either mask type: $W_i(Z_i = 1) = 1$, $W_i(Z_i = 0) = 0$
- Noncomplier under medical mask only: $W_i(Z_i = 1) = N$, $W_i(Z_i = 0) = 0$
- Noncomplier under cloth mask only: $W_i(Z_i = 1) = 1$, $W_i(Z_i = 0) = N$
- Noncomplier under either mask type: $W_i(Z_i = 1) = N$, $W_i(Z_i = 0) = N$

6) [3] Like the two-sided noncompliance case, we cannot proceed without additional assumptions. There may be different ways to go about this (you can formulate your own if you can think of any that work), but one possibility would be to assume that compliance is not dependent on type of mask - in other words, a person is either a complier (wears their mask at least 70% of hours) or a noncomplier (doesn't wear their mask at least 70% of hours), and this wouldn't change based on which type of mask he/she is assigned to wear. Do you think this assumption is plausible here? Why or why not? Can you think of a potential reason this assumption might be violated?

This may be plausible, if the wearer believes them both to be effective and if they are both equally comfortable. It may not be plausible if the health worker thinks one type of mask isn't effective so doesn't wear it, or if he/she finds one type of mask less comfortable so chooses to not wear it, but would have worn the other type.

7) [2] Under this assumption, how many compliance statuses are there? What are they?

Two; complier (wears assigned mask at least 70% of the time) and noncomplier (does not).

8) [2] Now we are back in more familiar territory! What additional assumption is needed in order to estimate the complier average causal effect? State the assumption by name and what it assumes.

Exclusion restriction for noncompliers: for noncompliers, potential outcomes do not depend on treatment assignment.

9) [2] Is this assumption plausible here? Why or why not? (We'll proceed under this assumption regardless of it's plausibility, but remember that the validity of our estimates depends on it!)

No! As long as the health worker wears their assigned mask *at all*, their potential outcome will be affected, even if they don't wear it at least 70% of the time. This means we have no way of separating the ITT for compliers and the ITT for noncompliers without additional data.

10) [3] Estimate the proportion of compliers. (Depending on how you defined W, note that using $\hat{ITT}_W$ may not work to estimate the proportion of compliers in this context.)

We can actually calculate this directly from the data, as compliance status here is fully observed. We can figure out the overall number of compliers based on the compliance proportion within each treatment group (given in the paper), and then divide this by the total.

```
n1 = 580
n0 = 569
p.c = (.566*n1 + .568*n0)/(n1+n0)
p.c
```

```
## [1] 0.5669904
```

11) [3] Use this to estimate the complier average causal effect. Interpret this in context.

```
itt.c = itt.y/p.c
itt.c
```

```
## [1] -0.03725453
```

However, note that this estimate depends on the exclusion restriction for noncompliers, which isn't viable here!

12) [3] Do you think the exclusion restriction for compliers is viable here? Why or why not? What do we gain by making this assumption?

Probably, but maybe not if health care workers have different views about the efficacy - maybe they will take other precautionary measures (such as more hand washing) if randomized to a cloth mask that they don't think is effective? The assumption is what lets the CACE be due to actually wearing the specific type of mask, not being assigned to the type of mask.

13) [3] Unlike the framework we explored in class, here we actually *can* directly observe compliance status for all units, under our assumption in (6). We aren't given the relevant data to do this, but might we estimate the complier average causal effect by looking at the difference in means for observed compliers? Why or why not?

Yes! We can actually observed compliance status for *everyone* so can directly estimate the CACE as the ITT for the complying units. Randomization should balance covariates within all subclasses, including within compliers.

14) [3] The paper includes "A post-hoc analysis adjusted for compliance and other potential confounders" with results in Table 4, but doesn't give any additional details so it's hard to tell what they actually did. Any guesses as to what they did?

Maybe two-stage least squares as a way to account for covariates and noncompliance.

15) [3] The study also included a control arm, where health care workers followed practice as usual, which for some meant wearing medical masks and for some meant wearing cloth masks. The study included a post-hoc analysis that compared the outcomes for all wearing medical masks (combining medical mask group and relevant controls) with all wearing cloth masks (comparing cloth mask group with relevant controls). What's wrong with this?

The control group is an observational study - people who choose to wear medical masks may differ than people who choose to wear cloth masks.

16) [2] What does this study tell you about the efficacy of cloth masks versus medical masks?

Cloth masks are less effective than medical masks as preventing infectious respiratory diseases. For this coronavirus crisis, we should prioritize obtaining/making more medical masks rather than relying on cloth masks for prevention.

Additional notes: Penetration of cloth masks by particles was almost 97%, as compared with 44% for medical masks. Moreover, the study authors point out that moisture retention, reuse of cloth masks, and poor filtration may actually result in increased risk of infection. They caution against the use of cloth masks as protection against respiratory infections.