ONLINE MATERIAL THAT ACCOMPANIES CHAPTER 12

Box 1. Representativeness, external validity, and randomized trials: two examples

Example 1: external validity and randomized trials

We are interested in whether a new weight-loss drug reduces obesity among school-age children in Farrlandia. We recruit children to be apart of a randomized trial of the drug. We require that children have a body mass index (BMI) above 25 but below 40 in order to participate in the trial. Further, since the drug is experimental, children with diabetes are excluded. Parents must be fully participatory in the trial, thus only parents who will commit to monitoring their children's drug regime closely and who commit to attending the study clinic once per week for educational classes on family nutrition and health are included. After children are recruited into the study, there is a one-month period where they are required to attend the clinic once per week for baseline surveys and measurements. Children who miss any appointments are deemed ineligible for the study in order to ensure that only participants who are likely to be adherent to study protocols are actually included in the study. After the one-month introductory period, children are then randomized to either receive the weight loss drug, or a placebo. They are then tracked for two years. Over the course of the study, the mean BMI in the group randomized to the drug declines from 31.5 to 26.7 (a difference of 4.8 points of BMI). The mean BMI in the group randomized to the placebo declines from 31.4 to 28.5 (a difference of 2.9 points of BMI). In sum, there was a reduction of 1.9 more points of BMI in the drug group than in the placebo group. The 95% confidence interval for this difference was 0.9 to 2.9, indicating that within reasonable bounds of chance in sampling variability, we could expect the difference to between a little as 0.9 or as large as 2.8 points of BMI.

Thus, the mean BMI declined more in the group randomized to the drug compared with the group randomized to placebo. Are these results externally valid to a broader population all overweight children in Farrlandia? What about overweight children in other places? What information would we need to know in order to inform this issue?

Example 2: external validity and representative samples

We are interested in whether a sales tax on sugar-sweetened beverages reduces obesity among children in Farrlandia aged 7 to 13. We enumerate all school-age children in Farrlandia, and take a random sample of 1,000 eligible Farrlandians (for eligibility, they only need to be between age 7 and 13 and living in Farrlandia). We measure their BMI before the tax goes into effect, and then measure their BMI after the tax across a two-year period. We find that the mean BMI of school-age children prior to the tax was 26.7 (95% C.I. 24.2, 29.3). Two years after the tax, the mean BMI of school-age children was 24.3 (95% C.I. 23.7, 24.9). We conclude that there is evidence that the tax lowered mean BMI among school-age children.

Another nearby town, Snowtown, is considering a similar tax given the success of the tax in lowering obesity in Farrlandia. Are the Farrlandian results externally valid to Snowtown? What information would we need to know about Farrlandians and Showtownians in order to inform this issue?

Back to example 1: external validity and the randomized trial

Let us return to example 1 and consider whether the study results of the weight loss drug would be applicable to the broader population of overweight children in Farrlandia. The first aspect we need to consider when making this assessment is whether we are confident in the result itself. A study result is only potentially externally valid to the extent that it is internally valid. In example 1 we carefully selected study participants who would adhere to study protocols, and we randomized the exposure. Within the bounds of sampling variability and chance, there is good reason to conclude that the result obtained is a good approximation of the causal effect of the drug for the population from which the study participants are drawn.

Given that we are confident about the internal validity of the results, next we need to consider the characteristics of the population from which the participants are drawn. These characteristics include those who are good adherers to the study protocol, do not have diabetes, and have parents who actively participate in the study process. Within this population, there is evidence that the drug is effective in reducing the population body mass index. Whether the result is externally valid beyond that specific underlying population of patients depends on whether the action of the drug interacts with any other factors, and whether these other factors have a different prevalence in the general patient population of overweight children who may be prescribed the drug. Importantly, some of the very eligibility criteria we used to increase internal validity might contribute to concerns about external validity. For example, depression might reasonably be hypothesized to be a component cause that interacts with the drug, i.e., co-occurs in the same sufficient cause through which the drug causes BMI lowering. In our study we excluded children who missed any early appointments. It is plausible that children with depression are more likely to be found in the missed appointment group than are children without depression. Therefore, it is plausible, perhaps even likely, that the prevalence of depression, a factor that potentially interacts with the drug in its effect of BMI, is different among children included in the study compared to children not included in this study. This would mean that our causal effect for the drug's influence on BMI may not be externally valid to populations with a different prevalence of depression than the study sample.

Back to example 2: external validity and the representative study

Let us return to example 2 and consider whether the study results would be applicable to Snowtown. The first aspect we need to consider when making this assessment is whether the result is internally valid. Thus, if the tax did not actually reduce rates of obesity, it will not be externally valid to Snowden no matter what the prevalence of causal partners of the tax are across the two populations. For example, if we learned that school lunches also changed to a healthier offerings during this time period, we would not want to make a causal claim that the tax reduced BMI, and certainly would not want to apply our result to Snowtown.

Assuming for the moment that the result is the causal effect of the tax on two-year BMI in Farrlandia, the answer to whether the result is externally valid in Snowtown requires us to consider what the potential causal partners of the soda tax may be. That is, under what conditions or exposures would a soda tax have a benefit on obesity? Let us suppose that one condition is the availability of soda. The tax may have less of an effect in places where sugar-sweetened beverages are difficult to find, such that the increase in price makes little difference in

terms of consumption since availability is low as is.

Suppose that soda is plentiful in Farrlandia, with multiple vending machines in every school and plenty of small grocery stores stocking soda throughout Farrlandian neighborhoods. In Snowtown, soda is hard to find, with no vending machines allowed in schools and only a few grocery stores stocking sodas. If soda availability is a component cause of soda tax, then the result in Farrlandia will not be valid in Snowtown, because availability differs between the two populations.

We provide an additional numerical example of a non-externally-valid sample results in Box 2 of the online material that accompanies Chapter 12.

Box 2. External validity and the prevalence of causal partners: an example

We are interested in the effect of a genetic factor on the incidence of high blood pressure. In Box 2, Figure 1, we conduct a study in three populations, sampling individuals who are initially free of high blood pressure, and following them forward in time for five years to estimate the incidence of high blood pressure. Each individual in the study provides a blood sample, which is genotyped. We will label those with the genetic sequence believed to be causally related to high blood pressure as exposed.

Box 2, Figure 1. Causal relationship between a genetic factor and high blood pressure

Sample 1 from Population 1

	Health indicator present	Health indicator absent	Total N
Exposed	50	50	100
Unexposed	200	200	400
Total N	250	250	500

$$\text{Risk difference} = \frac{50}{100} - \frac{200}{400} = 0.50 - 0.50 = 0\ [95\%\ \text{CI}\ (-0.11, 0.11)]$$

Sample 2 from Population 2

	Health indicator present	Health indicator absent	Total N
Exposed	100	100	200
Unexposed	50	250	300
Total N	150	350	500

$$\text{Risk difference} = \frac{100}{200} - \frac{50}{300} = 0.50 - 0.17 = 0.33\ [95\%\ \text{CI}\ (0.25, 0.41)]$$

Sample 3 from Population 3

	Health indicator present	Health indicator absent	Total N
Exposed	40	110	150
Unexposed	60	290	350
Total N	100	400	500

$$\text{Risk difference} = \frac{40}{150} - \frac{60}{350} = 0.27 - 0.17 = 0.10\ [95\%\ \text{CI}\ (0.01, 0.18)]$$

We find three different effect estimates across these three populations. In population 1, there is no association between the genetic factor and high blood pressure. In population 2, the genetic factor is associated with 33 additional cases of high blood pressure for every 100 people who have the genetic factor. In population 3, the genetic factor is associated with 10 additional cases of high blood pressure for every 100 people who have the genetic factor.

One reason that these three effect estimates may differ across populations is if there is a changing prevalence of a causal partner of the genetic factor. Suppose that the effect of having

the high-risk gene is activated when an individual is exposed to chronic levels of violence in their environment. Thus, in order to develop high blood pressure from the gene, one must also be exposed to violence.

The prevalence of violence exposure in Population 1 is 0%, the prevalence of violence exposure in Population 2 is 50%, and the prevalence of violence exposure in population 3 is 40%.

No one in Population 1 is exposed to violence. Therefore no one can develop high blood pressure from the genetic factor, because individuals need both the genetic factor and violence exposure to develop the disease from the genetic factor. Yet we see cases of high blood pressure in the exposed and unexposed. This is, of course, because there may be many other pathways to developing high blood pressure (e.g., smoking, poor diet, or a different genetic factor). Thus, individuals in the exposed group develop high blood pressure, but not because of the exposure. Because the exposure does not have its causal partner, the risk of high blood pressure is the same in the exposed and the unexposed.

Let us consider the exposed groups first. In Population 2, 50% of individuals are exposed to violence. Therefore, 50% of those exposed to the genetic factor will develop high blood pressure (in our sample, 100 out of 200). Similarly, in population 3, 40% of those who are exposed to the genetic factor will develop the disease because 40% of people are exposed to violence.

Now, let us consider the unexposed groups. In population 1, the risk of high blood pressure among those without the genetic factor is 17%. When we examine the risk of high blood pressure in Population 2 and Population 3, we can see that it is 17% for every population. Thus, the only difference across these populations is the prevalence of exposure to violence, and because exposure to violence interacts with the genetic factor, we see broad differences in the resulting magnitude of effect estimates.