ONLINE MATERIAL THAT ACCOMPANIES CHAPTER 10

Box 1. Natural and quasi-experimental experiments in observational epidemiology

In the typical randomized controlled trial, the investigator assigns the exposure randomly – some individuals are assigned to take the exposure, others are assigned to take the standard of care or a placebo. Because we cannot ethically assign individuals to become exposed to something that is harmful, many of the exposures that we believe to be the most influential cannot be evaluated in a randomized way. However, nature sometimes provides a way for us to collect observational data but analyze it with the rigor of a clinical trial.

For example, consider the Dutch Hunger Winter of 1944 - 1945. Due to a German blockade of food and fuel, several areas in the west of German-occupied Holland were severely malnourished over many months. Excellent records were kept of food supplies and rations, and the supplies and fuel were immediately restored after liberation. Thus, there is no reason to except there to be any systematic differences between the individuals who happened to be in the western area during the famine, versus before and after. While there was nothing natural about the food blockage to the Dutch cities, we consider studies of health indicators among those exposed compared to unexposed as natural experiments. This is why they are also sometimes referred to as quasi-experimental – while investigators did not assign the exposure status (which would make the study experimental), there can be a reasonable degree of confidence that the exposure is relatively randomly distributed within the population of interest. Findings from the Dutch Hunger Winter studies are considered to have a high level of validity in examining the effects of famine on health. Foundational research indicated, for example, that there was no difference in intelligence at age 18 among offspring of women who were pregnant and famine-exposed compared with offspring of women who were pregnant in cities without exposure to famine (Stein, Susser et al. 1972). These Dutch cohorts continue to be followed and continue to provide important information about the health effects of prenatal exposure to famine.

In general, any study which examines health outcomes among individuals who have been experienced an event or condition that was unexpected or outside of the individual's control are considered to be natural experiments, and inference from these types of studies can be made with a high level of rigor if indeed exposure was received in a relatively random manner. The goal, as in all epidemiologic studies, is to compare exposed and unexposed individuals who are comparable on all causes of the health indicator of interest other than the exposure. When we cannot randomize the exposure ourselves, we can still maintain a high level of rigor in our science by identifying these opportunities for study in situations in which individuals do not choose their exposure status.

Citations

Stein, Z., M. Susser, G. Saenger and F. Marolla (1972). "Nutrition and mental performance." Science 178(4062): 708-713.

Box 2. Limitations and design considerations in randomized studies

Equipoise and ethics in the randomized design

Many if not most exposures of interest to us for public health cannot be randomized because it is unethical or unfeasible. We cannot randomize individuals to exposures that we know or strongly suspect are harmful, or prevent them from receiving exposures that we know or suspect are beneficial. Thus, we cannot ask individuals to start smoking in order to study the health effects of smoking. Given that we already know that exercise is beneficial for many outcomes, we cannot ethically randomize people explicitly to refrain from exercise. We cannot prevent people from seeking treatment for alcoholism to know whether treatment for alcoholism is beneficial. In designing RCTs, individuals who do not receive the treatment that is being evaluated need to be offered the standard of care for the health outcome under study. For example, if a new treatment for alcoholism is under consideration, individuals who are not randomized to the treatment must be offered an appropriate standard of care for alcohol treatment. Further, there must be sufficient uncertainty about whether the factor of interest is potentially beneficial or harmful in order for an RCT to be ethical. Equipoise not only applies when the RCT is initially designed, but throughout the trial as well. The trial treatment must regularly be examined throughout the course of the study; if there is sufficient evidence that the treatment is outperforming or underperforming the standard of care, the trial must be halted and all study participants should be offered the most beneficial treatment.

Placebos

A critical and often problematic component of the randomized trial is what to provide the study participants who are not given the treatment of interest. As detailed above, we must ethically provide participants who do not receive the treatment with the current standard of care. In trials of pharmaceutical drugs, if there is no standard of care, participants who do not receive the treatment are often given a placebo drug. Placebo drugs are ineffective replacements designed to look and feel like the treatment of interest, but without the actual active ingredient. Placebos are critical, due to the well-documented placebo effect. That is, perceptions of symptoms as well as actual symptoms of the outcome change among people given an ineffectual treatment. Placebos often are designed to convince the participants that they have an active treatment; even common side effects of the treatment are replicated in the placebo drug so that participants believe they are being treated. Without a placebo, it would be impossible to know the extent to which the active ingredient in the treatment being tested is actually working, versus the effect of patients thinking that they are being treated and improving as a result.

Blinding

The role of the investigator and the study participants in understanding the treatment assignment is also critical – individuals who are randomized to the exercise condition and start to lose weight may also adopt healthier eating habits, for example. The investigators, in wanting to prove their hypothesis that exercise is beneficial for cardiovascular health, may look more carefully at the group who are not exercising and find more events in that group in order to prove their hypothesis. We mitigate the effects of potential bias on the part of the participant or the researcher by blinding when possible. That is, we do not tell patients whether or not they

are randomized to the active treatment group or the standard of care group. Further, we ensure that the investigators and study staff do not know the assignment of any particular participant. In many trials, it is also routine to blind the study statisticians to which groups actually received the treatment and which groups received the standard of care. Blinding is not always possible, however. For example, in our hypothetical exercise trial, it would be impossible to blind study participants to whether they are engaging in exercise. When possible, however, blinding ensures that no bias enters the trial with respect to the investigators, staff, and participants' knowledge of their treatment status.

Box 3. 95% confidence interval for a matched-pair odds ratio

Below we detail the steps to estimate a 95% confidence interval for a matched-pair odds ratio, and illustrate the steps using the hypothetical study data on fish oil and depression.

Step 1: Take the natural log of the matched-pair odds ratio

$$\text{Ln}(\text{Odds ratio}) = \text{Ln}(2.0) = 0.69$$

Step 2: estimate the standard error for the matched-pair odds ratio

$$\text{SE} = \sqrt{\frac{1}{\text{f}} + \frac{1}{\text{g}}}$$

$$\text{SE} = \sqrt{\frac{1}{10} + \frac{1}{5}} = 0.55$$

Step 3: Add and subtract the standard error multiplied by 1.96 from the log odds ratio

$$\text{Upper confidence interval, log odds} = \text{Ln}(\text{OR}) + 1.96 * \text{SE}[\text{Ln}[\text{OR}]]$$

$$\text{Lower confidence interval, log odds} = \text{Ln}(\text{OR}) - 1.96 * \text{SE}[\text{Ln}[\text{OR}]]$$

$$\text{Upper confidence interval, log odds} = 0.69 + 1.96 * 0.55 = 1.77$$

$$\text{Lower confidence interval, log odds} = 0.69 - 1.96 * 0.55 = -0.39$$

Step 4: Take the antilogarithm of the confidence interval bounds

$$\text{Upper confidence interval} = \text{e}^{1.77} = 5.87$$

$$\text{Lower confidence interval} = \text{e}^{-0.39} = 0.68$$

Step 5: interpret the results

In our study sample, individuals who consume low amounts of fish oil have 2.0 times the odds of depression compared with individuals who consume high amounts of fish oil. We are 95% confident that the odds ratio would be between 0.68 and 5.87 upon repeated sampling of the same underlying population.

Box 4. Is every variable associated with exposure and disease a potential source of non-comparability?

We have noted that one of the important steps in using stratification to control non-comparability is to assess whether the variable is associated with both exposure and outcome. Is every variable that is associated with exposure and outcome a potential source of non-comparability?

No. Even if a variable is associated with the exposure and the outcome, it may not necessarily contribute to non-comparability between the exposed and unexposed. There are two main ways in which variables would be associated with the exposure and the outcome but not contribute to non-comparability. First, if a variable is in a causal pathway between the exposure and the outcome that is part of the pathway of our hypothesis, it will be associated with the exposure and the outcome but it does not contribute to non-comparability. Second, if a variable is a consequence of both the exposure and the outcome, it would be associated with both exposure and disease but not contribute to non-comparability.

Example: factors in the causal pathway are not non-comparable variables

As an example, suppose that we are interested in whether exposure to tobacco smoke prenatally is a risk factor for offspring growth restriction during puberty and we hypothesize that part of the reason that prenatal tobacco smoke impacts growth restriction in puberty is through restricting the offspring birth weight (tobacco smoking during pregnancy is well-known to be causally associated with lower offspring birth weight). Thus, to be clear, the hypothesis is that prenatal exposure to tobacco causes low birth weight, and then this low birth weight causes growth restriction in puberty. In this case, we would not want to control for birth weight in examining the association between prenatal smoking and offspring birth weight. Let us consider what would happen if we were to control for birth weight through stratification of the analysis. Among those offspring with low birth weight, we would find that exposure to tobacco smoke was unrelated to offspring growth restriction, because we have restricted the analysis to those who have the intermediary outcome of interest, low birth weight. The same is true for those with normal birth weight – we would not find an association between the exposure and outcome, but not because there is no causal association – because we have restricted the analysis to those without the intermediary outcome.

In summary, factors that are on the causal pathway of interest between the exposure and outcome do not contribute to non-comparability. If we control for them, we will obstruct the ability to observe the true effects of the exposure on the outcome. Factors on the causal pathway of interest should not be controlled.

Example: factors that are consequences of exposure and outcome

Factors that are the consequence of, or result of, the exposure and the outcome will be associated with both exposure and outcome but do not contribute to non-comparability. An example of this can be found in one factor involved in the relation between post-menopausal hormone use and endometrial cancer: vaginal bleeding. Hormone use can cause vaginal bleeding, especially early in the protocol. Further, vaginal bleeding could be a symptom of (i.e., the result of) endometrial cancer. In this situation, in our data, vaginal bleeding would be associated with both the exposure (hormone use) and the outcome (endometrial cancer). But we would not want to control for vaginal bleeding in our analysis because vaginal bleeding is not contributing to non-comparability – it is not a factor that is an alternative explanation for any association we observe between hormone use and cancer because it is not a cause of cancer – rather, cancer is a cause of it. Thus, our analysis will be biased if we control for vaginal bleeding, but unbiased (on that factor at least) if we do not control for it.