**Distinguishing between Internal and External Validity.**

**Selection Bias vs. Generalizability**

Students often confuse selection bias with a lack of generalizability. Selection bias relates to the internal validity of the study, while generalizability speaks to the external validity of the study. Through these series of questions and examples below, you’ll practice learning to distinguish between the two.

**Readings:**

Vetter C, Devore EE, Wegrzyn LR, Massa J, Speizer FE, Kawachi I, Rosner B, Stampfer MJ, Schernhammer ES. Association between rotating night shift work and risk of coronary heart disease among women. J Ama. 2016;315(16):1726-1734. doi:10.1001/jama.2016.4454.Association.

Hatch EE, Hahn KA, Wise LA, Mikkelsen EM, Kumar R, Fox MP, Brooks DR, Riis AH, Sorensen HT, Rothman KJ. Evaluation of Selection Bias in an Internet-based Study of Pregnancy Planners. *Epidemiology*. 2016;27(1):98-104. doi:10.1097/EDE.0000000000000400.

**Before lab**, review the two articles listed above and answer questions 1-4 and 9-11 below. For our discussion we will use the term “study population” to refer to study participants.

1. In what way are selection bias and generalizability similar? How are they different?

*The are both about validity of the study. Selection bias is more about internal validity whereas generalizability is more about external validity.*

**The following questions are related to the article by Vetter et al.**

1. What is the main hypothesis of this paper?

Rotating night shift work is associated with coronary heart disease risk

1. What is the study population in this paper? Who are they and what kind of study design was used?

Prospective cohort study of 189,158 initially healthy

Women.

1. What do you think is the target population that the authors would like to generalize to?

Female nurses who work the night shift

1. Using Table 1 (p. 12), do you notice any obvious differences in age, sex, or race between the study population and the U.S population? Look at an external course to make this comparison, i.e 1990 Census data.

1. Briefly summarize the main results from this paper in a few sentences. Were the authors’ hypotheses supported?

1. Given the differences highlighted in question 5, would you expect that the results from this study could be applied to the entire U.S., for example? Why or why not? What type of population do you think this could be generalized to?

1. The article title states specifically that this study is among women. Why can’t these results be generalized to men? In other words, why might the results be different for a cohort that was more generalizable?

**The following questions are related to the article by Hatch et al.**

1. What is the main aim of this paper?

To evaluate selection bias in internet studies

1. How were participants enrolled in the Snart Gravid study?

4,801

1. How many women enrolled in the Snart Gravid study, and how does that compare to the number of women from the Danish Medical Birth Registry who were included after the exclusion criteria were met?

242,452 in the Danish medical registry

1. What issues do you think may arise with internet-based surveys? We’re specifically interested in what kind of person may be more likely to end up in this cohort, and whether the likelihood of inclusion in the cohort is related to the exposure and outcome of interest.

1. Using Table 2, were any RRs calculated from the Snart Gravid study different from RRs calculated using the Danish Birth Registry? If so, were they biased toward or away from the null compared to the birth registry results (the ‘true’ RRs)? We are looking for general trends by category.

1. Is there evidence of selection bias present in the Snart Gravid study? How did you determine this?

1. The RRs presented in Table 2 for the association between parity and pre-eclampsia are as follows: 1.7 (95% CI 1.3, 2.3) in Snart Gravid vs. 2.3 (95% CI 2.2, 2.4) in DMBR. Since we know that the RRs calculated from the Danish Medical Birth Registry are the ‘truth’, what can we say about the RR calculated from the Snart Gravid study? Is it biased toward or away from the null? How must the Snart Gravid cohort be different from the DMBR to generate this estimate (i.e. how must the prevalence of nulliparous women and/or pre-eclampsia be different)?

1. Do the authors discuss generalizability (or external validity) at all in this article? Note: in the abstract, the authors state “We cannot be certain that our results would apply to other associations or different populations.“ What do the authors mean by this statement?
2. Compare the methods used by the atuhors to assess selection bias in the Hatch et al. article with how you considered generalizability in the Vetter et al. article.

1. Do you see any potential issues with trying to apply the associations measured in this Danish registry based cohort to, hypothetically, pregnant women in all of Europe (i.e. generalizing to Europe)?