

# PHW250B Journal Club Assignment #1

## Causal inference and measures of disease

### Assignment Policies:

#### Formatting

*Each problem is followed by a box with a text field in which you can write your answer. Your answer should be contained within the box AND adhere to the max number of lines listed in the problem. Even if you have the ability, you should not remove page breaks, add pages or otherwise change the location of questions. This formatting makes the assignments easier for us to grade, meaning we can get them back to you sooner. On the first assignment there will be no penalty for improper formatting. In subsequent assignments (and exams!) only answers in the proper location are eligible for full credit.*

#### Group Work

*You may work with each other on problem sets. If you choose to do so, please submit your problem set individually but indicate below whether you worked with other students in the class (and list their names). You may also complete the problem set on your own.*

*Other Students with whom you worked (2 lines max):*

#### Grading

*We will grade up to three questions in each problem set (we will not announce which questions in advance).*

#### Solutions

*We will post detailed solutions to each problem set after all problem sets are submitted. It is your responsibility to compare your solutions to the ungraded problems to our posted solutions to ensure that you correctly answered the question.*

---

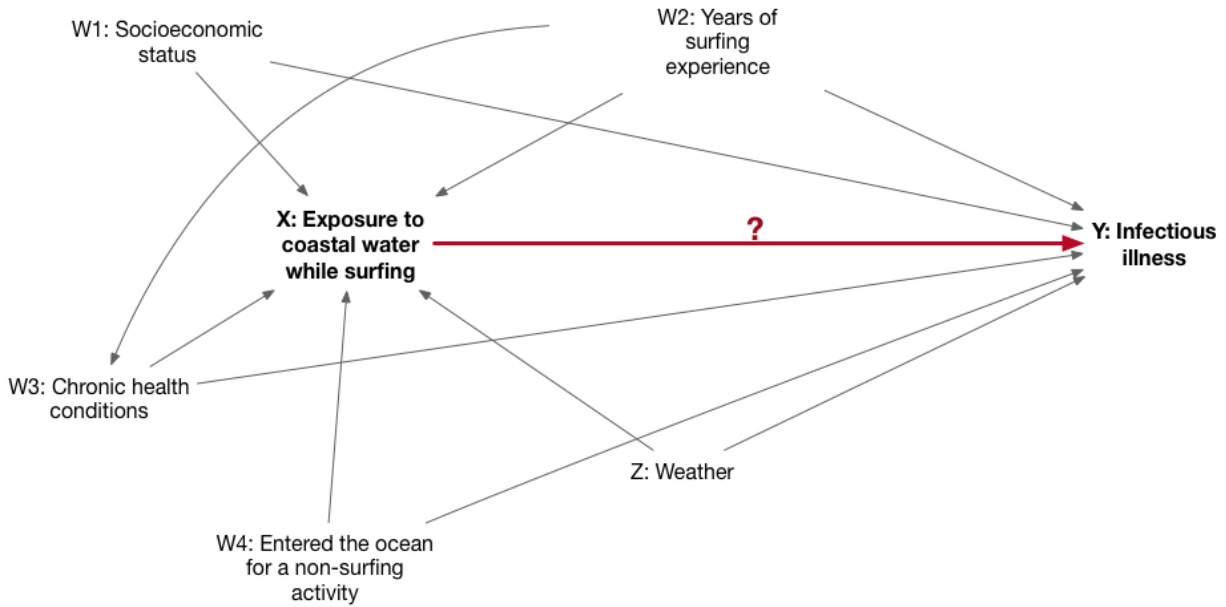
**Questions on this problem set are about Arnold et al., 2017 (only read pages 1-4)**

**Problem 1. What were the health outcomes of this study? (7 lines max)**

**Problem 2. Use information from the Introduction section of the paper to draw a DAG for this research question. Below, list the exposure and outcome, and variables that you believe are parents of the exposure and outcome in this study, including variables that were not measured in this study. Mark which variables were measured in this study. Use that information to complete the simple DAG below. (10 lines max)**

Fill in the exposure, outcome and at least one mentioned confounder into the DAG below.

**Problem 3. Define a structural causal model that corresponds to the simplified DAG below. (20 lines max)**



**Problem 4. Would you expect age, period, or cohort effects for this research question? Explain your answer briefly. (10 lines max)**

**Problem 5. What is a limitation of using prevalence instead of incidence in this study? (10 lines max)**

**Problem 6. What is an advantage of estimating incidence density instead of cumulative incidence in this study? (10 lines max)**

**Problem 7. Based on the description of the setting and study population, do you think this population was at a steady state? (10 lines max)**

Problem 8. Based on your answer to the previous question, write down the appropriate formula for calculating person-time in this study. In the formula fields here:

=

What assumption(s) does this formula **require when used to calculate incidence density**? (10 lines max)

**Problem 9A. The study reports rates of illness among surfers in San Diego, California. We want to compare those rates to the rates of illness among surfers in San Francisco, California. For the purpose of this question, assume that the amount and type of surfing is the same in both cities. Describe why we would use standardization to compare rates in these two cities. (10 lines max)**

**Question 9B. How would you go about performing direct standardization in this study?**  
**(20 lines max)**

**Question 9C. How would you go about performing indirect standardization? (20 lines max)**

**Problem 10. For practice using tables in these pdfs, please fill in the row and column totals in the two by two table below. Note that these values are per 1000 people, so there may be some non-whole numbers. Please round your answer to 2 decimal places.**

	Disease +	Disease -	Total
Exposure +	7.115	80.3	
Exposure -	8.234	500.1	
Total			