

Study Design

1 Test Recommendation System

You have created a recommendation system for movies in the data science course. Now you're starting a new job at Netflix. They would like to test whether using your recommendation system causes people to watch more movies in 2 months. You are given the task of planning a study to find out. Describe the ethical considerations you take, how you design the study, what data you collect, and how you analyze this data.

2 Study and Salary

You want to find out how things have gone with the students who have previously studied at the Department of Informatics. You are particularly interested in what in the study is responsible for how much people earn during the first five years after graduation. Does the choice of subject, grades, level of education (bachelor's or master's degree) matter? In addition, you will know if this depends on factors that have nothing to do with the education, such as gender or personality traits. To find out, the department has arranged an opportunity to send an email to all students who completed their studies during the last ten years. Make a questionnaire you want to send out, an analysis plan and discuss ethical considerations and possible problems you expect with the study.

3 Breast Cancer

Emma and Lucas are doing a cohort study on breast cancer. There are two common treatments for breast cancer. They try to find out which of the treatments more often leads to unsatisfactory outcomes. There are 100 patients who received treatment *A* and 100 patients who received treatment *B*. There were 37 patients who had unsatisfactory outcomes among the patients who received treatment *A* and 24 unsatisfactory outcomes among the patients who received treatment *B*. Now Emma and Lukas discuss how to interpret the results. Read through the discussion. How would you interpret the results? Explain it to Emma and Lukas. Also show an example of how the results can be misinterpreted.

Lukas: Treatment *A* worked poorly in 37% and treatment *B* in 24% of cases.

Then treatment B is $37/24 = 1.54$ times better than treatment A .

Emma: Obviously treatment B works better, we should stop using treatment A .

Lukas: It is true that treatment B is better in general, but there may be some groups where treatment A is best, for example among elderly patients or among patients with other diseases.

Emma: Then you are right. But we can probably say that for most patients, treatment B is better.

Lukas: In that case, we should only use treatment B until we have examined some patient groups in more detail.

Emma: But do we really know if the difference is causal?

Lukas: Even if it is not, it pays to switch everyone to treatment B