FMS1204S: Fraud, deception and data

Week 2

Types of medical data and misinterpretation

Common types of medical data

- In our readings for next week we will continue to discuss dishonesty with data in the medical world.
- Some of the most fascinating instances of misleading with data occur innocently through misunderstandings of data collection and interpretation.
- It is helpful to understand common types of medical data:
  - Data from non-interventional studies — observational studies;
  - Data from interventional studies — designed experiments..

Non-interventional studies

- Prospective study
  - We are studying a certain disease say. A group of individuals without the disease is obtained and characteristics which may be risk factors for the disease (exposure variables) are measured.
  - The group is followed forward in time and we look for differences in the exposure variables between the participants in the study who develop the disease and those who don't.
- Some disadvantages
  - Time consuming and expensive.
  - Large study populations are required, and it may be hard to generalize from the study population to the general population.
  - It may take a long time for enough disease cases to arise to allow for sensible analysis.

Running example - smoking and lung cancer

- In the 1950's Richard Doll and A. Bradford Hill reported a prospective study involving 50,000 medical doctors in which they investigated the association between smoking and lung cancer.
- 50,000 doctors (without lung cancer at the beginning of the study) were interviewed about their smoking habits and health habits and followed for five years.
- This study and similar prospective studies established a relationship between the likelihood of developing lung cancer and how much someone smokes.

Non-interventional studies (cont.)

- Retrospective study
  - A group of individuals with the disease is identified. Then we attempt to match the diseased individuals with similar disease free individuals (similar in terms of characteristics like age, sex, etc.)
  - Look at whether there are differences in exposure variables between the diseased and disease free individuals.
- Some disadvantages
  - It may be difficult to find participants for the disease free controls matching the disease cases.
  - To measure the exposure variables we rely on the recall and honesty of the participants.
  - Prone to selection bias (in taking a sample of disease cases, for example, those who die quickly from the disease may not be properly represented).

Running example - smoking and lung cancer

- Previously we mentioned a propsective study of smoking and lung cancer by the researchers Hill and Doll.
- Prior to this prospective study they had published in 1952 a retrospective study in the British Medical Journal.
- They located several hundred lung cancer patients, and matched them to similar patients (in terms of age, sex, socioeconomic status) without lung cancer.
- Among the lung cancer patients there were about 10 times as many smokers.

Non-Interventional studies (cont.)

- Often we are interested in inferring a cause and effect relationship. That is, I want to know if exposure to some risk factor will increase my risk of getting a disease.
- Cause and effect is difficult to infer from observational (non-interventional) studies. This is because of the possible presence of confounding variables - an apparent relationship between disease and exposure may be caused by a confounder that is associated with both disease status and exposure.

Running example - smoking and lung cancer

- It took a long time for the cause and effect relationship between smoking and lung cancer to be established. Why?
- If the evidence comes only from observational (non-interventional) studies, then there is the possibility of confounding.
- For example, it is possible that there might be some genetic basis to the preference to smoke, and that this same genetic factor could cause lung cancer. If this were the case, then stopping a person with a genetic predisposition to smoke from smoking would not alter their risk of getting lung cancer.

Running example - smoking and lung cancer

- The case for a causal effect was considered proven once there were a range of retrospective and prospective studies pointing in the same direction, when it was established that as the amount of exposure (smoking) increased then the risk of lung cancer increased, and when plausible biological mechanisms for the effect were understood.

Interventional studies

- The "gold standard" for proving cause and effect is a randomized controlled double blind study. Here the participants are assigned to two (or more) treatments or exposures with the choice of treatment chosen at random.
- The random assignment to treatment ensures that any differences between the groups can only be due to the treatment.
- Ideally both the researcher and the participant should be unaware of what treatment was administered (double blinding).
- Often such studies cannot be conducted for ethical reasons (for instance we don’t expose people knowingly to some possibly toxic quantity). For this reason, for example, it would be unethical to do a randomized controlled double blind study to prove that smoking causes lung cancer.

Misinterpretations

- It is no exaggeration to say that most articles of a medical nature in the newspapers are misleading.
- Typically the report will involve discussion of a single study and its conclusions.
- The problems of the design of the study are rarely appreciated or communicated.
- Frequently the writer will not have the time or expertise to put the reported study in the context of many other studies on the same topic that may have different conclusions.
- If a number of different studies of different types point in the same direction then we can be more confident that a reported effect is real.

Groups and readings for next week

Group 1

- Goldacre, Ben (2008). *Bad Science*, Fourth Estate, London, pp. 207–224 (Chapter 12, How the media promote the public misunderstanding of science).
- Questions to address specifically from the reading:
  - Why the media would promote the public misunderstanding of science?

Group 2

- Ellison, George (2006). Medicine in black and white: BiDil: race and the limits of evidence-based medicine. Significance 3 (3), 118–121.
- This reading discusses a clinical trial in which a heart drug was found to be ineffective, but after the end of the trial the researchers were able to find a subgroup of the patients that appeared to benefit. Issues to discuss specifically:
  - Explain the meaning of the term personalized medicine.
  - Do you think the data from the trial was used appropriately? Do you think the analysis of the data was misleading, and do you think this was innocent or intentional?

Group 3

- Goldacre, Ben (2008). *Bad Science*, Fourth Estate, London, pp. 28–62 (Chapter 16, The Media's MMR Hoax).
- Questions to address specifically from the reading:
  - What concerns were raised about the MMR vaccine?
  - On what data were the concerns based? What were the limitations of the data?
  - Do you think that Dr Andrew Wakefield behaved ethically in raising concerns about the MMR vaccine?

Group 4

- Rosenthal, J. (2005). Struck by lightning: The curious world of probabilities. London: Granta Publications, pp. 234-246 (Chapter 16, Ignorance, Chaos, and Quantum Mechanics).
- Questions to address specifically from the reading:
  - Do you believe truly randomness? What do you feel about randomness? What do you feel about ”determinism”?

Group 5

- Weir, Chris and Murray Gordon (2011). Fraud in clinical trials: Detecting it and preventing it. Significance 8 (4), 164–168.
- Clinical trials are used to investigate whether new drugs and treatments are effective and safe. However, fraud does occur.
  - What are the incentives for researchers committing fraud in clinical trials?
  - How might fraud be detected? How might it be prevented?