Lesson Plan

- Review of Experimental Design
- Prospective and Retrospective studies
- Ethics of Experimentation
- Government Rules
Experimental Design

- Run over time to see endpoint
- Matched? not randomized
- Select a sample without risk factor
- Select a sample with risk factor
- Prospective study - difficult to run, costly in both money and time
- Typically say two levels for risk factor, two levels for outcome
- Association between them?
- Risk factor, clinical outcome or endpoint
Case-control studies

Longitudinal studies
together

Cross-sectional studies (infrequent) - risk and outcome sampled

Bias in inference

are samples representative of population with outcome

less expensive, can over-sample rare outcomes

for what risk factors were present

Retrograde study - samples determined from outcomes, look back

Experimental design cont.
Screening test, widely used, e.g., mammography, AIDS, etc.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Test+</th>
<th>Test-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>000</td>
<td>950</td>
</tr>
<tr>
<td>No Disease</td>
<td>1000</td>
<td>10</td>
</tr>
<tr>
<td>False negative</td>
<td>False positive</td>
<td>Disease incidence</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>0.00001</td>
<td>0.010</td>
<td>1/100</td>
</tr>
<tr>
<td>0.0025</td>
<td>0.0677</td>
<td>1/200</td>
</tr>
<tr>
<td>0.00005</td>
<td>0.084</td>
<td>1/500</td>
</tr>
<tr>
<td>0.0005</td>
<td>0.00913</td>
<td>1/1000</td>
</tr>
<tr>
<td>0.0001</td>
<td>0.01666</td>
<td>1/10000</td>
</tr>
<tr>
<td>0.000</td>
<td>0.016664</td>
<td>1/100000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>False negative</th>
<th>False positive</th>
<th>Disease incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0001</td>
<td>0.016664</td>
<td>1/1000000</td>
</tr>
</tbody>
</table>
A related problem - Berkson's fallacy

An example - Is there a connection between incidence of locomotor disease and respiratory disease?

<table>
<thead>
<tr>
<th></th>
<th>Loc Disease pres</th>
<th>Resp Disease pres</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hospital subsample</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>2376</td>
<td>184</td>
</tr>
<tr>
<td>207</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td><strong>General population</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.06</td>
<td>219</td>
<td>17</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

- Berkson’s fallacy
Hospitalization rates differ according to what patient presents. In fact:

- Chance of hospitalization given no loc disease and no resp disease is 0.09.
- Chance of hospitalization given no loc disease and resp disease is 0.7.
- Chance of hospitalization given loc disease and no resp disease is 0.07.
- Chance of hospitalization given loc disease and resp disease is 0.29.

What's going on?
2. Ethics of Experimentation

Scientists in almost every field have codified rules or guidelines that determine the ethical limits of the research they can perform. These principles may be enforced by:

- Individual good sense.
- Universities.
- Government or funding agencies.

Corporations have fewer restrictions than government or academic researchers.

Professional societies must interact with scientists from every field. We need to understand the rules that apply in different situations.
Ethical principles - Nuremberg Code

- Voluntary consent
- Anticipates scientific benefits
- Benefits outweigh risks
- Avoid suffering
- Do animal experiments first
- No intentional death or disablement
- Protection from harm
- Subjects free to stop
- Investigator will stop if harm occurs
- Conducted by a qualified investigator
- Subjects free to stop
2.1 Review Procedures

The National Institute of Health requires that all grantees must have their research approved by an Institutional Review Board (IRB). Most universities have extended the scope of their IRBs to cover all human and animal research conducted by faculty, staff, and students. At Duke, there are two IRBs—one for the medical school, and one for all other departments.

Several of the statistics faculty are involved in projects that require IRB approval. For example, I am involved with the Center for Geospatial Medicine (CGM) - Outcomes like premature birth, low birth weight, neural tube defects; "Causes" such as exposure, psycho-social stressors, genetic predispositions; Spatial pattern.
Another example: experiments to determine ways to improve student learning and retention. It is likely that members of this class will be invited to participate in these studies, and later we shall discuss the approval process for such experiments.
Medical clinical trials are highly regulated and especially important. Human lives are on the line, and it is important to distinguish good therapies from harmful therapies as quickly as possible. Many statisticians are involved in this kind of work, employed by both the pharmaceutical industries and the Food and Drug Administration, which regulates new treatments. Examples of problems are:

- Unanticipated side effects; e.g., a blood pressure drug that causes early so that the trial can be modified.
- Differences in efficacy on subgroups; e.g., a drug that helps women but harms men is a pattern that should be noticed early so that the trial can be modified.

In cases involving clinical trials, there is an additional requirement for an Interim Monitoring Board, which periodically looks at the unblinded data to see whether there are emerging problems. Examples of problems are:

- Unanticipated differences in efficacy on subgroups.
- Unanticipated side effects.

Liver damage.

Medical clinical trials are highly regulated and especially important.
The establishment of IRBs and other systems for professional ethics was largely a response to three famous experiments in the 1960s and early 1970s.

1. Stanley Milgram's psychology experiments on obedience to authority, known as the "Abraham Experiment," in 1960-1963.

These experiments were viewed (perhaps incorrectly) as unethical.
In the main version of Milgram's experiment, subjects responded to an ad in the New Haven newspaper which promised a small sum of money in exchange for participation in a psychology experiment. When the subject arrived, he met an actor who pretended to be the examiner. The Actor/Learner was given a list of pairs of nonsense syllables and was instructed to spend 30 minutes memorizing the pairs. The actor/learner was chosen to be the Learner and the subject was chosen to be the Teacher. The experiment needed two people, and through a fake coin toss the actor and another respondent to the ad, and the Scientist, said the another respondent to the ad, and the Scientist, said the Scientist said the

2.2.1 Obedience to Authority
The Scientist explained that the purpose of the experiment was to determine how far the subject would go on the orders of the Scientist.

The point of the experiment was to determine whether a small amount of pain, in the form of an electric shock, would encourage better recall of the nonsense pairs.

Both the Learner and the Teacher were given a shock of 15 volts on the lowest setting (15 volts) of the machine.

If the Learner/Actor FAILED, the Teacher/Subject was supposed to provide the matching syllable.

If the Learner/Actor FAILED, the Teacher/Subject was supposed to read one of the nonsense pairs, and the Teacher/Subject was supposed to administer a shock of ever increasing intensity. The scale on the machine ran from 15 volts to 450 volts, and at 360 volts the machine dial was marked DANGER.

The Scientist explained that the purpose of the experiment was to determine how far the subject would go on the orders of the Scientist.
The Learner/Actor went to another room, was attached to the electrodes, and the Teacher/Subject and the Scientist returned to the room with the control panel for the shock generator. The Learner/Actor communicated his answers by intercom.

The Learner/Actor pretended to do a poor job. As the voltage increased, he pretended to answer further questions. At 300 volts, he pounded the wall frantically, and after 315 volts there was no sound from the room. Usually, the Teacher/Subject wants to stop. But the Scientist calmly says (according to a strict script) that the experiment should proceed.

Sometimes, the Teacher/Subject exhibits great stress, but in many cases follows the instructions of the Scientist. Often the subject exhibits great stress, but in many cases follows the script (according to a strict script) that the experiment should proceed.

At 360 volts, the Learner/Actor refused to answer further questions. At 380 volts, he pounded the wall. Then he began to scream, to beg them to stop, and finally shouted that he refused to answer further questions. At 400 volts, the Learner/Actor sobbed. After 415 volts, there was no sound from the room. But the Scientist calmly says (according to a strict script) that the experiment should proceed.

At 450 volts, the Learner/Actor refused to answer further questions. At 470 volts, he pounded the wall. Then he began to scream, to beg them to stop, and finally shouted that he refused to answer further questions. At 500 volts, the Learner/Actor sobbed. After 515 volts, there was no sound from the room. But the Scientist calmly says (according to a strict script) that the experiment should proceed.

At 550 volts, the Learner/Actor refused to answer further questions. At 570 volts, he pounded the wall. Then he began to scream, to beg them to stop, and finally shouted that he refused to answer further questions. At 600 volts, the Learner/Actor sobbed. After 615 volts, there was no sound from the room. But the Scientist calmly says (according to a strict script) that the experiment should proceed.

At 650 volts, the Learner/Actor refused to answer further questions. At 670 volts, he pounded the wall. Then he began to scream, to beg them to stop, and finally shouted that he refused to answer further questions. At 700 volts, the Learner/Actor sobbed. After 715 volts, there was no sound from the room. But the Scientist calmly says (according to a strict script) that the experiment should proceed.

At 750 volts, the Learner/Actor refused to answer further questions. At 770 volts, he pounded the wall. Then he began to scream, to beg them to stop, and finally shouted that he refused to answer further questions. At 800 volts, the Learner/Actor sobbed. After 815 volts, there was no sound from the room. But the Scientist calmly says (according to a strict script) that the experiment should proceed.

At 850 volts, the Learner/Actor refused to answer further questions. At 870 volts, he pounded the wall. Then he began to scream, to beg them to stop, and finally shouted that he refused to answer further questions. At 900 volts, the Learner/Actor sobbed. After 915 volts, there was no sound from the room. But the Scientist calmly says (according to a strict script) that the experiment should proceed.

At 950 volts, the Learner/Actor refused to answer further questions. At 970 volts, he pounded the wall. Then he began to scream, to beg them to stop, and finally shouted that he refused to answer further questions. At 1000 volts, the Learner/Actor sobbed. After 1015 volts, there was no sound from the room. But the Scientist calmly says (according to a strict script) that the experiment should proceed.
The upshot is that about 72% of men, and about 63% of women, went to the top of the dial. And a much larger percentage went very far on the dial before refusing to proceed. Therefore, the ethical question was whether it was appropriate to perform this kind of experiment, and whether it was morally equivalent to Nazis.

Some of the subjects were deeply upset by the realization that they were morally equivalent to Nazis.

After the experiment, the subjects were debriefed. They were told that their behavior was quite normal, and that they should not take this too seriously. Some of were deeply upset by the realization that they were morally equivalent to Nazis.

The subjects experienced great distress, and none were deeply upset. Some of were morally equivalent to Nazis.
Laud Humphreys was a graduate student at Harvard who wanted to study the demographic characteristics of mail homosexuals. At the time, the sociology community generally thought that homosexuals were more likely than the general population to have low intelligence, criminal records, menial jobs, and/or foreign-born. He became friendly with the community, so that he could be sure that the subjects were gay, and even acted as “Watch Queen” to warn when the police entered the parking lot.

Humphreys went to highway rest stations where gay men would meet. He took down the license plate numbers of the cars that parked there. He became friendly with the community, so that he could be sure that the subjects were gay, and even acted as “Watch Queen” to warn when the police entered the parking lot. He took down the license plate numbers of the cars that parked there.

Humphreys went to highway rest stations where gay men would meet. He became friendly with the community, so that he could be sure that the subjects were gay, and even acted as “Watch Queen” to warn when the police entered the parking lot. He took down the license plate numbers of the cars that parked there.
Humphreys took his license plate numbers to the Division of Motor Vehicles and paid a small fee to get the registered address. (The DMV no longer provides that service.)

Humphreys kept the name list in a locked safe, to which only he and the major professor had the combination. After the experiment was complete, he destroyed it.

Humphreys' major professor was performing a public health survey for the U.S. government. With the professor's permission, Humphreys sent the survey form to his list of subjects, so that he could compare their responses to those of a random sample of the community.
Humphreys’ experiment found that gays were not different from respondents in any significant way, aside from the fact that slightly fewer respondents were married and they had slightly fewer children. However, the sociology community was disturbed by his methodology. His actions as a researcher violated his role as Watch Queen may have created a social obligation that could have led to the subjects being “outed.” His list of names posed a danger—under unlikely contingencies, it could have led to the subjects being “outed.” His list of names posed a danger—under unlikely contingencies, it could have led to the subjects being “outed.” He had not obtained informed consent from his subjects for their participation. He had not obtained informed consent from his subjects for their participation. However, the sociology community was disturbed by his methodology. His role as Watch Queen may have created a social obligation that could have led to the subjects being “outed.” This exploded a key misconception about homosexuality.
The Stanford Prison Experiment

In 1971, Zimbardo developed an experiment in which 24 Stanford students were paid $15/day to play roles as guards and prisoners. They were selected from among 70 potential recruits through a psychological screening system designed to weed out social misfits. From those who qualified, 8 were randomly chosen to be prisoners and 16 were chosen to be guards. Details are at www.childdevelopment.com/stanford/prison.htm.

Then there was an elaborate simulation of arrest, delousing, imprisonment,
The experiment was supposed to last two weeks. But within only six days, and a few days after, many of the prisoners had broken down and wanted to be released. And many of the guards had been broken down and wanted to be released. And many of the prisoners had broken down and wanted to be released. And

Zimbardo and his colleagues were also affected. They began thinking like prison wardens, trying to keep the prisoner subjects in detention and making the experiment seem less onerous. They manipulated the visiting hours so as to make the experiment seem less onerous. They were revolted by what they saw.

Christina Maslach, a recent Stanford Ph.D., joined the team after the experiment was in progress. She was revolted by what she saw, and insisted that they terminate the experiment immediately. The experiment ended after only six days.
Zimbardo wrote: “We had learned through videotapes that the guards
were escalating their abuse of prisoners in the middle of the night when
their boredom had driven them to over more pornographic and
degrading abuse of the prisoners.”

Their boredom had driven them to escalate their abuse of prisoners by
thinking no researchers were watching and the experiment was
surrounded the Milgram and Humphreys studies. But the issues raised
by this experiment are obvious.

Zimbardo’s early termination obviated much of the ethical criticism;
and that normal people behaved badly. The Stanford Prison Experiment showed very convincingly (though not
statistically) that people adopted social roles according to situational
expectations. And that those roles could easily take over, so that normal
people behaved badly. By this experiment these issues are 
obvious.
In 1974, partly in response to these experiments, the National Research Act created a commission to identify basic ethical principles in human subjects research. The result of that commission is known as the Belmont Report. The report established three central principles:

- Respect for Individuals
- Beneficence
- Justice

These principles are supposed to be guidelines, and IRBs may, in principle, approve a research proposal that does not accord with them.
Respect for Individuals implies that:

1. Individuals should be treated as autonomous agents.
2. People of diminished autonomy are entitled to protection.
3. Prisoners, children, people in impoverished countries cannot participate except in exceptionally innocuous cases. (But these are often difficult.)

In practice, this means that:

- Thereshouldbeinformedconsent. (But this is often difficult.)
- Thereshouldbenocoercion. (But what about fair compensation?)
- Praiseworthy, children, people in impoverished countries cannot participate except in exceptionally innocuous cases. (But these are often difficult.)

(But what about fair compensation?)
Beneficence implies that:

1. One should do no harm. (But few experimenters have such intention.)

2. Maximize possible benefits, minimize possible harms.

Many clinical trials appear problematic. Double-blinding gets tricky if one must perform invasive surgery. A cost-benefit analysis, using proper statistical risk analysis, is probably a better way to express this. In practice, the first principle is often said but the second is often used.
Justice implies that the benefits, burdens, and risks of experiments should be shared fairly.

This is unclear. Some hold that benefits should be shared according to individual effort, some say they should be shared according to societal contribution, and some say that should be shared according to merit.

For example, in a new drug therapy, you may prefer to wait for the outcome of the clinical trial, and receive the best therapy, rather than participate in the trial. In that case you have unfairly benefitted from the risks that others have taken.
There are other forces which affect statistical experiments. For example, the Office of Management and Budget (OMB) must approve every survey that a federal employee gives to more than 9 people.

The OMB wants to:

- minimize burden on citizen respondents
- ensure adequate response rates
- ensure good survey design
- verify that questions are appropriate and useful (and not politically problematic)

There are other forces which affect statistical experiments.