Putting your life in the hands of a coin: Randomized trials.

Steve Simon
P.Mean Consulting
www.pmean.com

2. Why do I offer this webinar for free?

I offer free statistics webinars partly for fun and partly to build up goodwill for my consulting business,

- www.pmean.com/consult.html.

Also see my Facebook and LinkedIn pages

- www.facebook.com/pmean
- www.linkedin.com/in/pmean

I provide a free newsletter about Statistics, The Monthly Mean.

- www.pmean.com/news
- www.facebook.com/group.php?gid=302778306676

3. Abstract

- In research studies that compare a treatment group and a control group, you need to assess whether the comparison is a fair comparison—an apples to apples comparison. Randomization is a simple method that insures that patients assigned to the treatment group are comparable to patients assigned to the control group. There are, however, practical and ethical constraints that can sometimes prevent the use of randomization.

4. Objectives

In this class you will learn how to:

- describe how covariate imbalance can threaten the validity of a research study,
- explain how randomization prevents covariate imbalance, and
- understand the practical and ethical limitations to randomized studies.

5. Sources

Part of the material for this webinar comes from:

- Simon SD. Statistical Evidence in Medical Trials, What Do the Data Really Tell Us? 2006. Oxford University Press: Oxford, England.
- Simon SD. Is the randomized clinical trial the gold standard of research?. J Androl. 2001 Nov-Dec;22(6):938-43.
- Stats #32a: Statistical Evidence: Apples or Oranges? Randomized studies.
  - http://www.childrens-mercy.org/stats/training/hand32a.asp

6. Pop quiz #1

When the demographic profile of the patients in your treatment group differ sharply from the profile of patients in your control group, you have:

1. covariate imbalance,
2. observational data,
3. response bias,
4. spectrum bias,
5. stratified data,
6. don't know/not sure

7. Pop quiz #2

Randomization is not practical:

1. when doctors believe that the new treatment is superior to the current standard
2. when patients have a strong preference for a particular treatment
3. when the experiment requires deliberate exposure of patients to something that is known to be harmful
4. randomization is impractical for all of the above situations
5. randomization can be applied easily in all of the above situations
6. don't know/not sure

8. Pop quiz #3

The following approaches are credible alternatives to randomization:

1. alternating between treatment and control
2. assigning all new patients to the treatment group and choosing controls from a medical database
3. assigning treatment group on the basis of the last digit of your birthday
4. letting the doctor choose whether a patient gets into the treatment group or the control group
5. none of these approaches is as effective as randomization
6. don't know/not sure

9. Covariate imbalance

Almost all research involves comparison. Do women who take Tamoxifen have a lower rate of breast cancer recurrence than women who take a placebo? Do left-handed people die at an earlier age than right-handed people? Are men with severe vertex balding more likely to develop heart disease than men with no balding?

10. Covariate imbalance

When you make a comparison between a treatment group and a control group, you want a fair comparison. You want the control group to be identical to the treatment group in all respects, except for the treatment in question. You want an apples-to-apples comparison.

11. Covariate imbalance

Sometimes, however, you get an unfair comparison, an apples-to-oranges comparison. The control group differs on some important characteristics that might influence the outcome measure. This is known as covariate imbalance. Covariate imbalance is not an insurmountable problem, but it does make a study less authoritative.

12. Covariate imbalance

Women who take oral contraceptives appear to have a higher risk of cervical cancer. But covariate imbalance might be producing an artificial rise in cancer rates for this group. Women who take oral contraceptives behave, as a group, differently than other women.

13. Covariate imbalance

For example, women who take oral contraceptives have a larger number of pap smears. This is probably because these women visit their doctors more regularly in order to get their prescriptions refilled and therefore have more opportunities to be offered a pap smear. This difference could lead to an increase in the number of detected cancer cases. Perhaps the other women have just as much cancer, but it is more likely to remain undetected.

14. Covariate imbalance

- There are many other variables that influence the development of cervical cancer: age of first intercourse, number of sexual partners, use of condoms, and smoking habits. If women who take oral contraceptives differ in any of these lifestyle factors, then that might also produce a difference in cervical cancer rates.
- The possibility that oral contraceptives causes an increase in the risk of cervical cancer is quite complex; a good summary of all the issues involved is available at:
  - www.jhuccp.org/pr/a9/a9chap5.shtml.

15. Randomization

One way to avoid most of the problems with imbalanced covariates is to use randomization. Randomization is the assignment of treatment groups through the use of a random device, like the flip of a coin or the roll of a die, or numbers randomly generated by a computer. Randomization is not always possible, practical, or ethical. But when you can use randomization, it greatly adds to the credibility of the research study.

16. Randomization

In a randomized study, the researchers have a high degree of control over the patients. They decide who gets what. This is a hallmark of a randomized design and it only can occur when the patients and/or their doctors have no say in the assignment. This is an incredible gift that patients in a research study offer you. They sacrifice their ability to choose between two therapies and instead let that choice be decided by the flip of a coin.

17. Randomization

Randomization helps ensure that both measurable and immeasurable factors are balanced out across both the standard and the new therapy, assuring a fair comparison. Used correctly, it also guarantees that no conscious or subconscious efforts were used to allocate subjects in a biased way.

18. Randomization

Randomization relies on the law of large numbers. With small sample sizes, covariate imbalance may still sneak in. A study examining the probability of covariate imbalance (Hsu 1989) showed that total sample sizes less than 10 could have a 50% chance or higher of having a categorical covariate with levels twice as large in one group than the other. This study also showed that total sample sizes of 40 or greater would have very little chance of such a serious imbalance.

19. A fishy story about randomization

I was told this story but have no way of verifying its accuracy. It is one of those stories that if it is not true, it should be. A long, long, time ago, a research group wanted to examine a pollutant to find concentration levels that would kill fish. This research required that 100 fish be separated into five tanks, each of which would get a different level of the pollutant. The researchers caught the first 20 fish and put them in the first tank, then put the next 20 fish in a second tank, and so forth. The last 20 fish went into the fifth tank. Each fish tank got a different concentration of the pollutant.

20. A fishy story about randomization

When the research was done, the mortality was related not to the dosage, but to the order in which the tanks were filled, with the worst outcomes being in the first tank filled and the best outcomes in the last tank filled. What happened was that the slow-moving, easy-to-catch fish (the weakest and most sickly) were all allocated to the first tank. The fast-moving, hard-to-catch fish (the strongest and healthiest) ended up in the last tank.

.

21. Concealed allocation

Another important aspect of randomization is concealed allocation, which is withholding the randomization list from those involved with recruiting subjects. This concealment occurs until after subjects agree to participate and the recruiter determines that the patient is eligible for the study. Only then is a sealed envelope opened that reveals the treatment status. Concealed allocation can also be done through a special phone number that the doctor calls to discover the treatment status.

22. Concealed allocation

If the randomization list is not concealed, doctors have the ability to consciously or unconsciously influence the composition of the groups. They can do this by applying exclusion criteria differentially or by delaying entry of a certain healthier (or unhealthier) subject so he/she gets into the ‘desirable’ group. Unblinded allocation schemes tend, on average to overstate the effectiveness of the new therapy by 30–40% (Schulz 1996).

23. Ethical and practical constraints on randomization

There are many situations where randomization is not practical or possible. Sometimes patients have a strong preference for one particular treatment and would not consider the possibility of being randomized into a different treatment. Surgery is one area with strong patient preferences especially for newer approaches like laparoscopic surgery (Lefering 2003).

24. Ethical and practical constraints on randomization

Randomization is also problematic for interventions that are already known to be effective. While further research would help better define these advantages, you cannot ask half of your patients to sacrifice the benefits of the new intervention.

25. Ethical and practical constraints on randomization

Randomization also does not work when you are studying noxious agents, like second-hand cigarette smoke or noisy workplaces. It would be unethical to deliberately expose people to any of these agents, so we have use non-randomized studies of people who are unfortunate enough to be trapped in settings with noxious agents.

26. Ethical and practical constraints on randomization

Sometimes researchers just do not want to go to the effort of randomizing. If you assign the treatment or therapy, rather than letting the patients and their doctors choose, you have to expend a lot of energy. Is it worth the effort? It is usually faster and cheaper to use existing nonrandomized databases. You get a lot larger sample size for your money. Depending on the situation, that might be enough to counterbalance the advantages of randomization.

27. Variations on randomization

There are three variations to randomization where the researchers have control over treatment assignment, but they use something other than a table of random numbers for the assignment. The first approach, minimization, is a credible and reasonable choice, but the other two approaches, alternating assignment and haphazard assignment, do not have much to recommend them.

28. Variations on randomization

An alternative, when the researchers have sufficient control, is to allocate the assignments so that at each step, the covariate imbalance is minimized.

So if the treatment group has a slight surplus of older patients and the next patient to join the study is also older than average, then that patient would be assigned to the control group so as to reduce the age discrepancy.

29. Variations on randomization

Another approach used in place of randomization is to alternate the assignment, so that every even patient is in the treatment group and every odd patient is in the control group. Alternating assignment was popular in trials before World War II; it was felt that researchers would not understand and not tolerate randomization (Yoshioka 1998).

30. Variations on randomization

Alternating assignment seems on the surface to be a good approach, but it can sometimes lead to trouble. This is especially true when one patient has a direct or indirect influence on the next patient. You may have seen this level of influence if you grow vegetables in a garden. If you have a row of cabbages, for example, you will often see a pattern of big cabbage, little cabbage, big cabbage, little cabbage, etc.

31. Variations on randomization

What happens, if the cabbages are planted a bit too closely, is that one of the cabbages will grow just a bit faster at first. It will extend into the neighboring cabbage's territory, stealing some of the nutrients and water, and thus growing even faster at the expense of the neighbor. If you assigned a fertilizer to every other cabbage, you would probably see an artificial difference because of the alternating pattern in growth within a row.

32. Variations on randomization

Haphazard assignment uses some arbitrary value like a birthdate or social security number to assign patients to groups. Often it is the evenness/oddness of the arbitrary number that determines the treatment assignment. For example, patients born on even-numbered dates would be assigned to the treatment group and those born on odd-numbered dates would be assigned to the control group. An arbitrary or haphazard number is never going to be as good as a purely random number. The haphazard assignment will always cast a shadow of doubt over the research study.

33. Practice exercises

For each of the following abstracts, randomization was NOT used. Explain why it would be impractical or unethical to conduct a randomized experiment in each of these settings.

1. Body fatness during childhood and adolescence and incidence of breast cancer in premenopausal women: a prospective cohort study. Heather J Baer, Graham A Colditz, Bernard Rosner, Karin B Michels, Janet W Rich-Edwards, David J Hunter and Walter C Willett. Breast Cancer Research 2005, 7:R314-R325 doi:10.1186/bcr998.

Introduction Body mass index (BMI) during adulthood is inversely related to the incidence of premenopausal breast cancer, but the role of body fatness earlier in life is less clear. We examined prospectively the relation between body fatness during childhood and adolescence and the incidence of breast cancer in premenopausal women. **Methods** Participants were 109,267 premenopausal women in the Nurses' Health Study II who recalled their body fatness at ages 5, 10 and 20 years using a validated 9-level figure drawing. Over 12 years of follow up, 1318 incident cases of breast cancer were identified. Cox proportional hazards regression was used to compute relative risks (RRs) and 95% confidence intervals (CIs) for body fatness at each age and for average childhood (ages 5–10 years) and adolescent (ages 10–20 years) fatness. **Results** Body fatness at each age was inversely associated with premenopausal breast cancer incidence; the multivariate RRs were 0.48 (95% CI 0.35–0.55) and 0.57 (95% CI 0.39–0.83) for the most overweight compared with the most lean in childhood and adolescence, respectively (P for trend < 0.0001). The association for childhood body fatness was only slightly attenuated after adjustment for later BMI, with a multivariate RR of 0.52 (95% CI 0.38–0.71) for the most overweight compared with the most lean (P for trend = 0.001). Adjustment for menstrual cycle characteristics had little impact on the association. **Conclusion** Greater body fatness during childhood and adolescence is associated with reduced incidence of premenopausal breast cancer, independent of adult BMI and menstrual cycle characteristics. http://breast-cancer-research.com/content/7/3/R314

2. **Impact of a nurses' protocol-directed weaning procedure on outcomes in patients undergoing mechanical ventilation for longer than 48 hours: a prospective cohort study with a matched historical control group**. Jean-Marie Tonnelier, Gwenaël Prat, Grégoire Le Gal, Christophe Gut-Gobert, Anne Renault, Jean-Michel Boles and Erwan L'Her. Critical Care 2005, 9:R83-R89 doi:10.1186/cc3030. **Introduction** The aim of the study was to determine whether the use of a nurses' protocol-directed weaning procedure, based on the French intensive care society (SRLF) consensus recommendations, was associated with reductions in the duration of mechanical ventilation and intensive care unit (ICU) length of stay in patients requiring more than 48 hours of mechanical ventilation. **Methods** This prospective study was conducted in a university hospital ICU from January 2002 through to February 2003. A total of 104 patients who had been ventilated for more than 48 hours and were weaned from mechanical ventilation using a nurses' protocol-directed procedure (cases) were compared with a 1:1 matched historical control group who underwent conventional physician-directed weaning (between 1999 and 2001). Duration of ventilation and length of ICU stay, rate of unsuccessful extubation and rate of ventilator-associated pneumonia were compared between cases and controls. **Results** The duration of mechanical ventilation (16.6 ± 13 days versus 22.5 ± 21 days; P = 0.02) and ICU length of stay (21.6 ± 14.3 days versus 27.6 ± 21.7 days; P = 0.02) were lower among patients who underwent the nurses' protocol-directed weaning than among control individuals. Ventilator-associated pneumonia, ventilator discontinuation failure rates and ICU mortality were similar between the two groups. **Discussion** Application of the nurses' protocol-directed weaning procedure described here is safe and promotes significant outcome benefits in patients who require more than 48 hours of mechanical ventilation. http://ccforum.com/content/9/2/R83

3. **Extravascular lung water in patients with severe sepsis: a prospective cohort study**. Greg S Martin, Stephanie Eaton, Meredith Mealer and Marc Moss. Critical Care 2005, 9:R74-R82 doi:10.1186/cc3025. **Introduction** Few investigations have prospectively examined extravascular lung water (EVLW) in patients with severe sepsis. We sought to determine whether EVLW may contribute to lung injury in these patients by quantifying the relationship of EVLW to parameters of lung injury, to determine the effects of chronic alcohol abuse on EVLW, and to determine whether EVLW may be a useful tool in the diagnosis of acute respiratory distress syndrome (ARDS). **Methods** The present prospective cohort study was conducted in consecutive patients with severe sepsis from a medical intensive care unit in an urban university teaching hospital. In each patient, transpulmonary thermodilution was used to measure cardiovascular hemodynamics and EVLW for 7 days via an arterial catheter placed within 72 hours of meeting criteria for severe sepsis. **Results** A total of 29 patients were studied. Twenty-five of the 29 patients (86%) were mechanically ventilated, 15 of the 29 patients (52%) developed ARDS, and overall 28-day mortality was 41%. Eight out of 14 patients (57%) with non-ARDS severe sepsis had high EVLW with significantly greater hypoxemia than did those patient with low EVLW (mean arterial oxygen tension/fractional inspired oxygen ratio 230.7 ± 36.1 mmHg versus 341.2 ± 92.8 mmHg; $P < 0.001$). Four out of 15 patients with severe sepsis with ARDS maintained a low EVLW and had better 28-day survival than did ARDS patients with high EVLW (100% versus 36%; $P = 0.03$). ARDS patients with a history of chronic alcohol abuse had greater EVLW than did nonalcoholic patients (19.9 ml/kg versus 8.7 ml/kg; $P < 0.0001$). The arterial oxygen tension/fractional inspired oxygen ratio, lung injury score, and chest radiograph scores correlated with EVLW ($r2 = 0.27$, $r2 = 0.18$, and $r2 = 0.28$, respectively; all $P < 0.0001$). **Conclusions** More than half of the patients with severe sepsis but without ARDS had increased EVLW, possibly representing subclinical lung injury. Chronic alcohol abuse was associated with increased EVLW, whereas lower EVLW was associated with survival. EVLW correlated moderately with the severity of lung injury but did not account for all respiratory derangements. EVLW may improve both risk stratification and management of patients with severe sepsis. http://ccforum.com/content/9/2/R74

4. **Breast implants following mastectomy in women with early-stage breast cancer: prevalence and impact on survival**. Gem M Le, Cynthia D O'Malley, Sally L Glaser, Charles F Lynch, Janet L Stanford, Theresa HM Keegan and Dee W West. Breast Cancer Res 2005, 7:R184-R193 doi:10.1186/bcr974. **Background** Few studies have examined the effect of breast implants after mastectomy on long-term survival in breast cancer patients, despite growing public health concern over potential long-term adverse health effects. **Methods** We analyzed data from the Surveillance, Epidemiology and End Results Breast Implant Surveillance Study conducted in San Francisco–Oakland, in Seattle–Puget Sound, and in Iowa. This population-based, retrospective cohort included women younger than 65 years when diagnosed with early or unstaged first primary breast cancer between 1983 and 1989, treated with mastectomy. The women were followed for a median of 12.4 years (n = 4968). Breast implant usage was validated by medical record review. Cox proportional hazards models were used to estimate hazard rate ratios for survival time until death due to breast cancer or other causes for women with and without breast implants, adjusted for relevant patient and tumor characteristics. **Results** Twenty percent of cases received postmastectomy breast implants, with silicone gel-filled implants comprising the most common type. Patients with implants were younger and more likely to have in situ disease than patients not receiving implants. Risks of breast cancer mortality (hazard ratio, 0.54; 95% confidence interval, 0.43–0.67) and nonbreast cancer mortality (hazard ratio, 0.59; 95% confidence interval, 0.41–0.85) were lower in patients with implants than in those patients without implants, following adjustment for age and year of diagnosis, race/ethnicity, stage, tumor grade, histology, and radiation therapy. Implant type did not appear to influence long-term survival. **Conclusions** In a large, population-representative sample, breast implants following mastectomy do not appear to confer any survival disadvantage following early-stage breast cancer in women younger than 65 years old. http://breast-cancer-research.com/content/7/2/R184

38. Conclusion

Randomization is the use of a random device to assign patients to a treatment group or control group. When the sample size is sufficiently large, randomization prevents covariate imbalance in your experiment. Randomization is not practical if patients have a strong preference for a particular treatment and is unethical if it forces some patients to endure a harmful exposure.

39. Pop quiz #1

When the demographic profile of the patients in your treatment group differ sharply from the profile of patients in your control group, you have:

1. covariate imbalance,
2. observational data,
3. response bias,
4. spectrum bias,
5. stratified data,
6. don't know/not sure

40. Pop quiz #2

Randomization is not practical:

1. when doctors believe that the new treatment is superior to the current standard
2. when patients have a strong preference for a particular treatment
3. when the experiment requires deliberate exposure of patients to something that is known to be harmful
4. randomization is impractical for all of the above situations
5. randomization can be applied easily in all of the above situations
6. don't know/not sure

41. Pop quiz #3

The following approaches are credible alternatives to randomization:

1. alternating between treatment and control
2. assigning all new patients to the treatment group and choosing controls from a medical database
3. assigning treatment group on the basis of the last digit of your birthday
4. letting the doctor choose whether a patient gets into the treatment group or the control group
5. none of these approaches is as effective as randomization
6. don't know/not sure