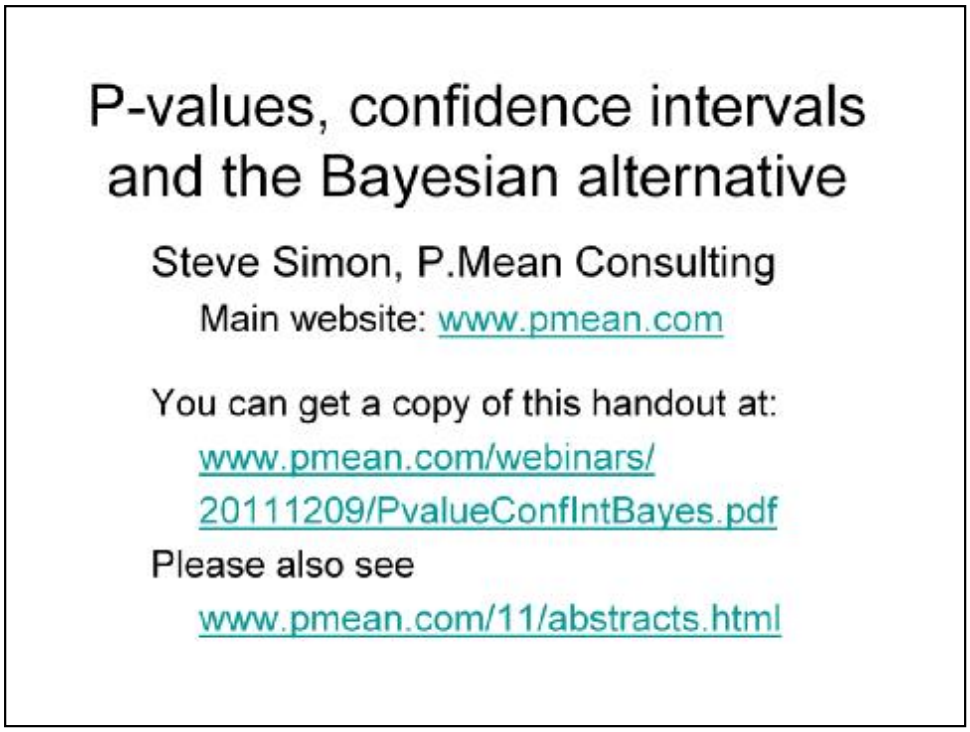


I provide a lot of information about Statistics for free. Start at my main website, but also take a look at my email newsletter, The Monthly Mean.

Facebook: www.facebook.com/pmean

LinkedIn: www.linkedin.com/in/pmean

Main website: www.pmean.com

Newsletter: www.pmean.com/news

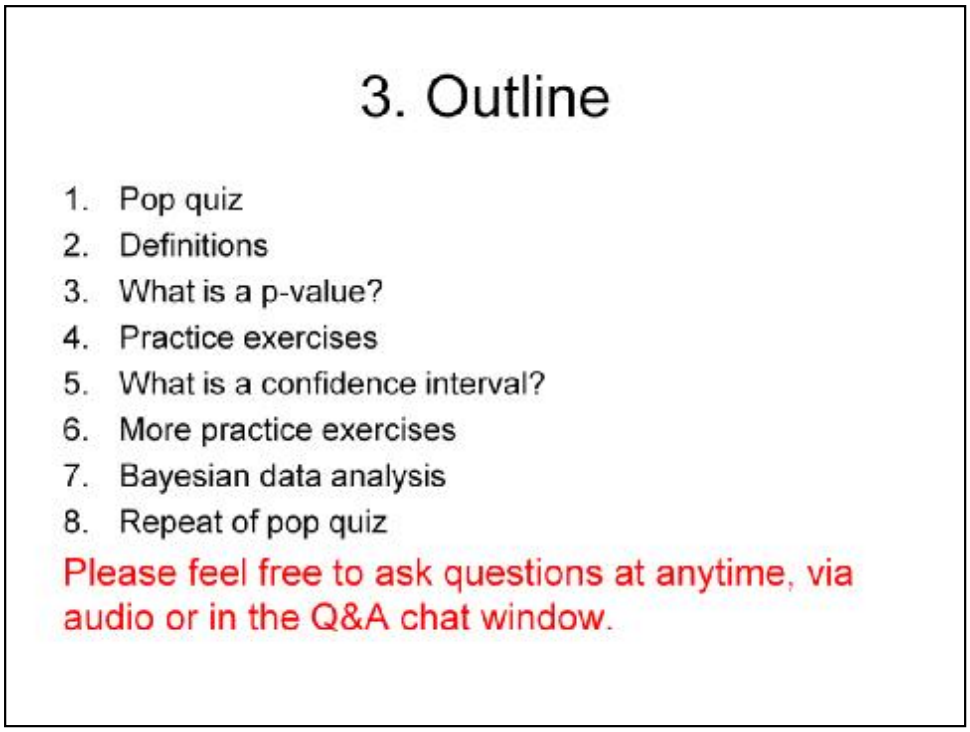
Twitter: http://twitter.com/profmean

Please note that this webinar is heavily dependent on computer technology that frequently fails to deliver what it promises. I've had just about everything bad happen to me during these webinars, and it helps to just laugh it off and carry on as best as you can. If you're having problems with the audio or video, please let me know. There may not be a fix, but there is usually a work-around.

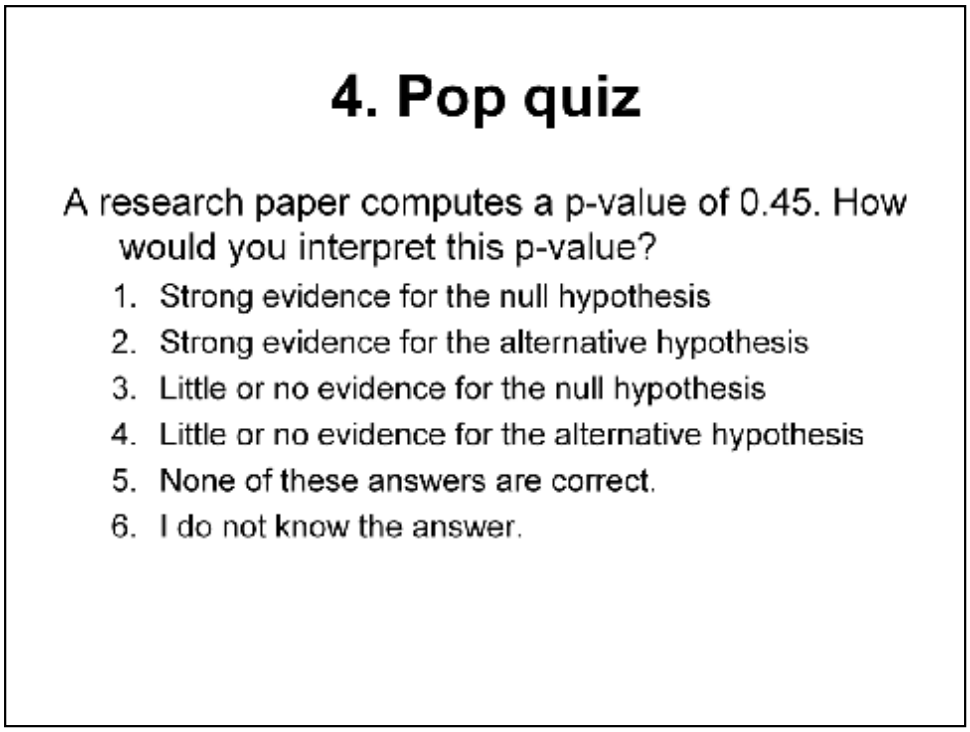
2. Learning objectives

- In this seminar, you will learn how to:
  - distinguish between statistical significance and clinical significance;
  - define and interpret p-values;
  - describe the advantages of confidence intervals;
  - explain the difference between informative and diffuse priors;

P-values are the fundamental tools used in most inferential data analyses. They are possibly the most commonly reported statistics in the medical literature. Unfortunately, p-values are subject to frequent misinterpretations. In this presentation, you will learn the proper interpretation of p-values, and the common abuses and misconceptions about these statistics.



The pop quiz will not be graded. I do need to lay out some definitions. You may already know these terms, but it is still helpful to review them to assure that we are communicating using a common set of assumptions. I'll guide you through the first practice exercise, and then ask you to respond to the next two in the chat window. Then I'll repeat the pop quiz at the end. Please feel free to ask questions at anytime. I plan to finish no later than two hours after I have started, but will stick around if there are any questions.



Don’t worry if you don’t know the answer to this right away. I’m just trying to get a sense of how much people know in advance of the seminar. I’ll repeat this pop quiz at the end as an informal evaluation of whether I have been effective at communicating my message to you.

5. Pop quiz

A research paper computes a confidence interval for a relative risk of 0.82 to 3.94. What does this confidence interval tell you.

1. The result is statistically significant and clinically important.
2. The result is not statistically significant, but is clinically important.
3. The result is statistically significant, but not clinically important.
4. The result is not statistically significant, and not clinically important.
5. The result is ambiguous.
6. I do not know the answer.

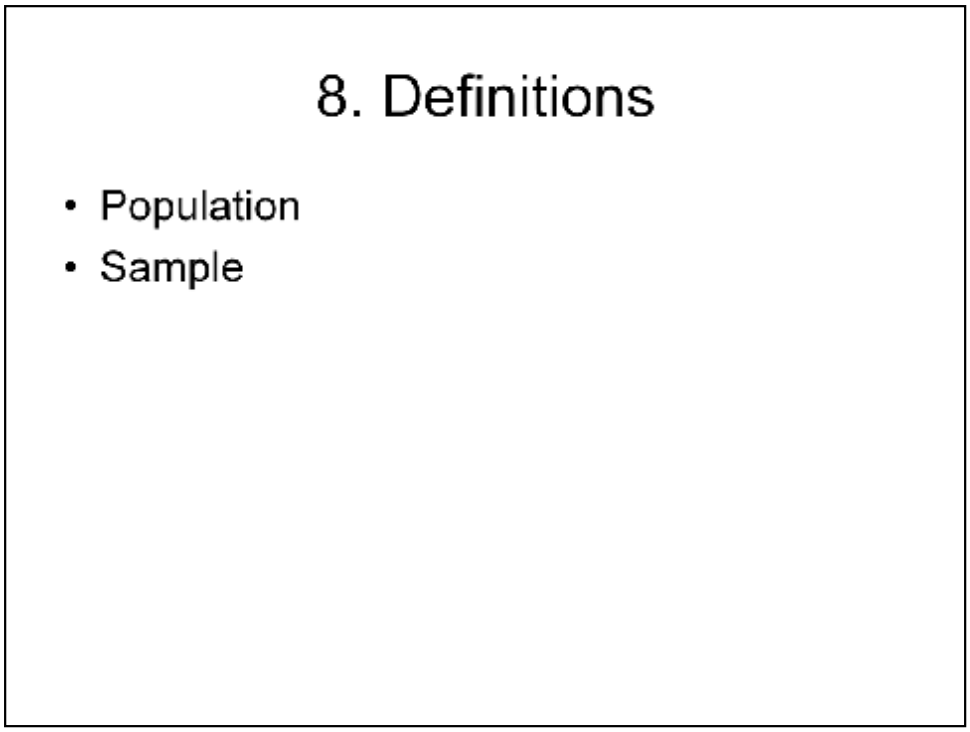
6. Pop quiz

A Bayesian data analysis can incorporate subjective opinions through the use of

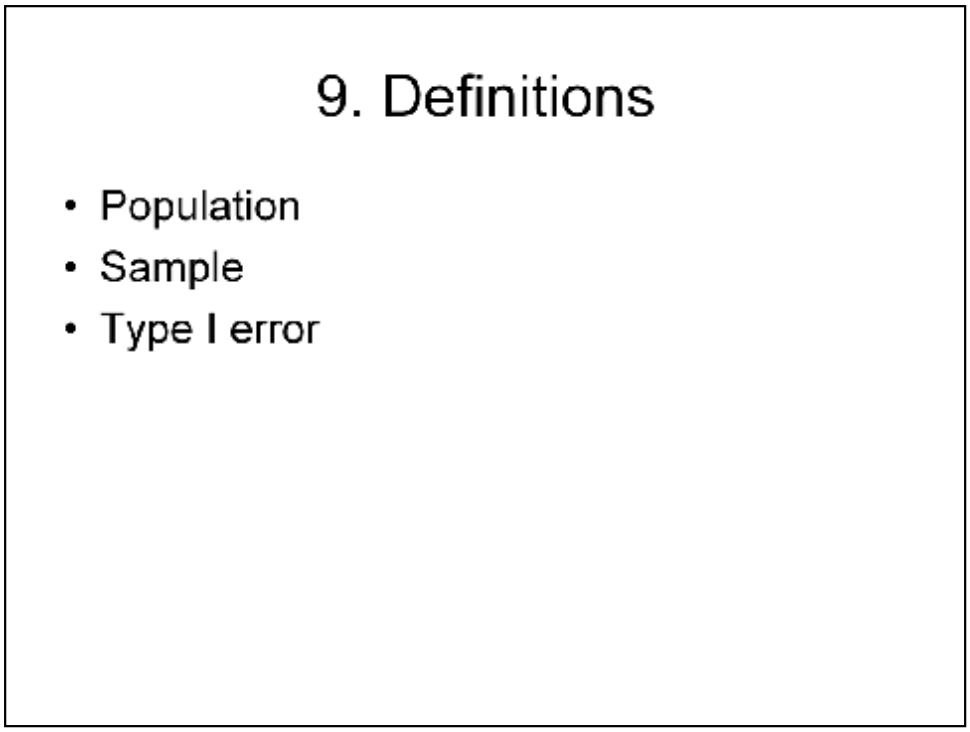
1. Bayes rule.
2. data shrinkage.
3. a prior distribution.
4. a posterior distribution.
5. p-values.
6. I do not know the answer.

A population is a collection of items of interest in research. The population represents a group that you wish to generalize your research to. Populations are often **defined in terms of demography, geography, occupation, time, care requirements, diagnosis, or some combination of the above**.

An example of a population would be all infants born in the state of Missouri during the 1995 calendar year who have one or more visits to the Emergency room during their first year of life.

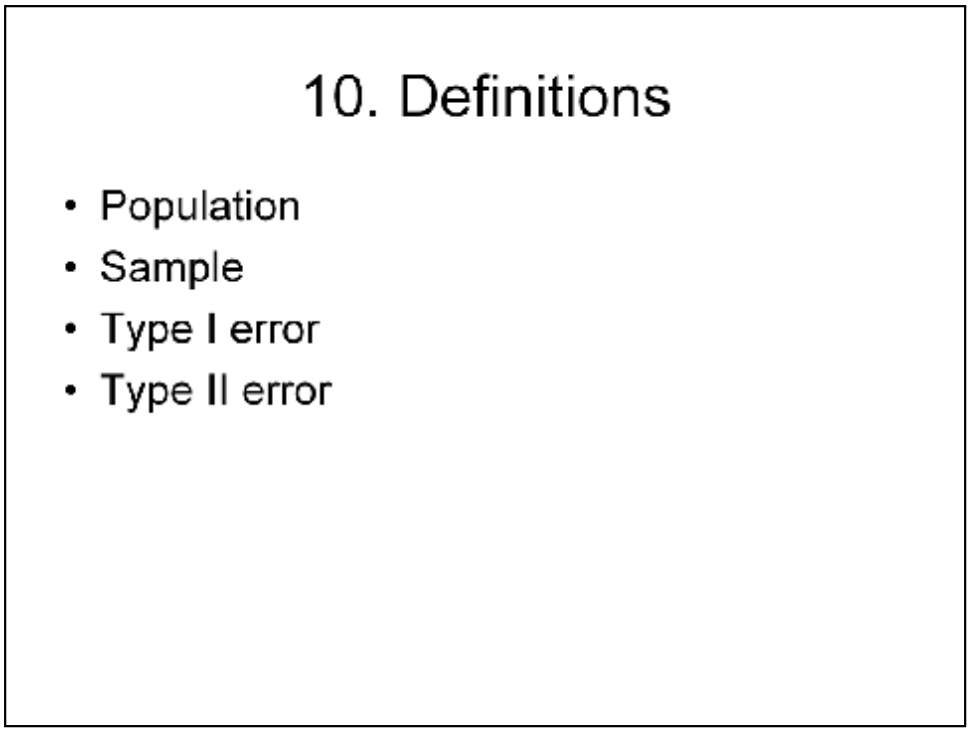


A sample is a subset of a population. A random sample is a subset where every item in the population has the same probability of being in the sample. **Usually, the size of the sample is much less than the size of the population**. The primary goal of much research is to use information collected from a sample to try to characterize a certain population.



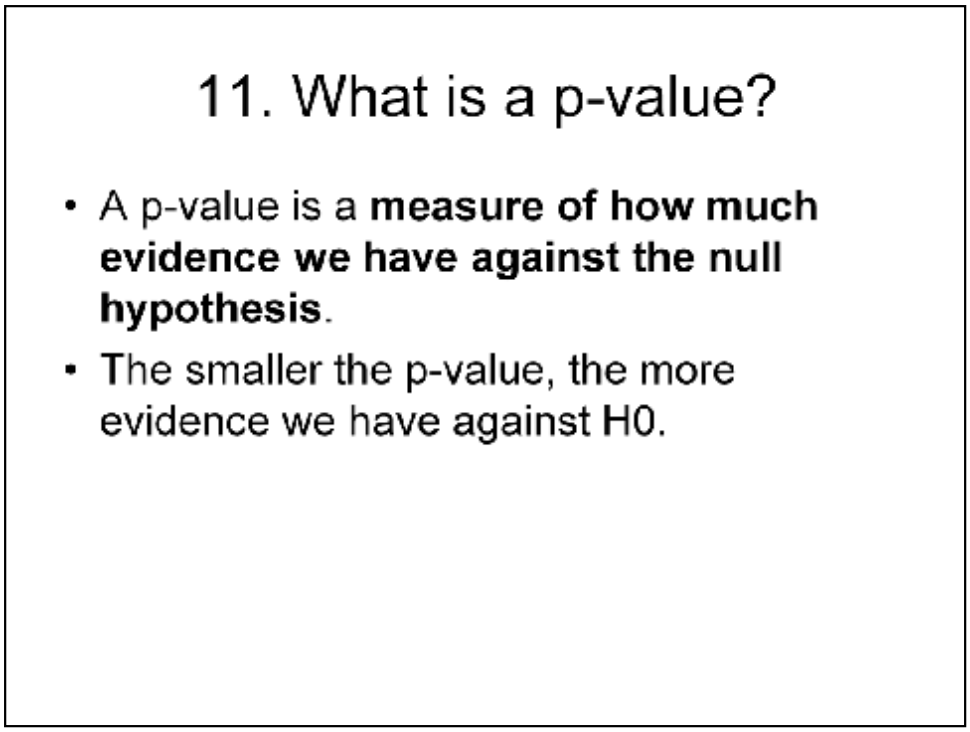
In your research, you specify a null hypothesis (typically labeled H0) and an alternative hypothesis (typically labeled Ha, or sometimes H1). By tradition, the null hypothesis corresponds to no change. A Type I error is **rejecting the null hypothesis when the null hypothesis is true**.

Example: Consider a new drug that we will put on the market if we can show that it is better than a placebo. In this context, H0 would represent the hypothesis that the average improvement (or perhaps the probability of improvement) among all patients taking the new drug is equal to the average improvement (probability of improvement) among all patients taking the placebo. A Type I error would be **allowing an ineffective drug onto the market**.



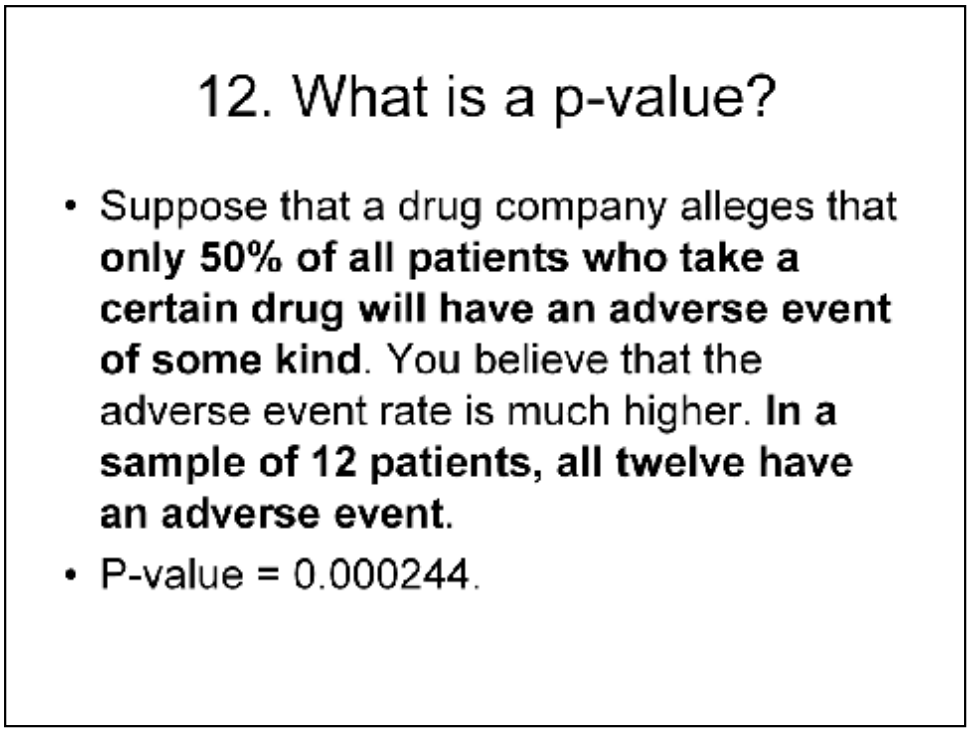
A Type II error is **accepting the null hypothesis when the null hypothesis is false**. **Many studies have small sample sizes that make it difficult to reject the null hypothesis,** even when there is a big change in the data. In these situations, a Type II error might be a possible explanation for the negative study results.

Example: Consider a new drug that we will put on the market if we can show that it is better than a placebo. In this context, H0 would represent the hypothesis that the average improvement (or perhaps the probability of improvement) among all patients taking the new drug is equal to the average improvement (probability of improvement) among all patients taking the placebo. A Type II error would be **keeping an effective drug off the market**.



The null hypothesis, traditionally represented by the symbol H0, represents the hypothesis of no change or no effect. The smaller the p-value, the more evidence we have against H0.

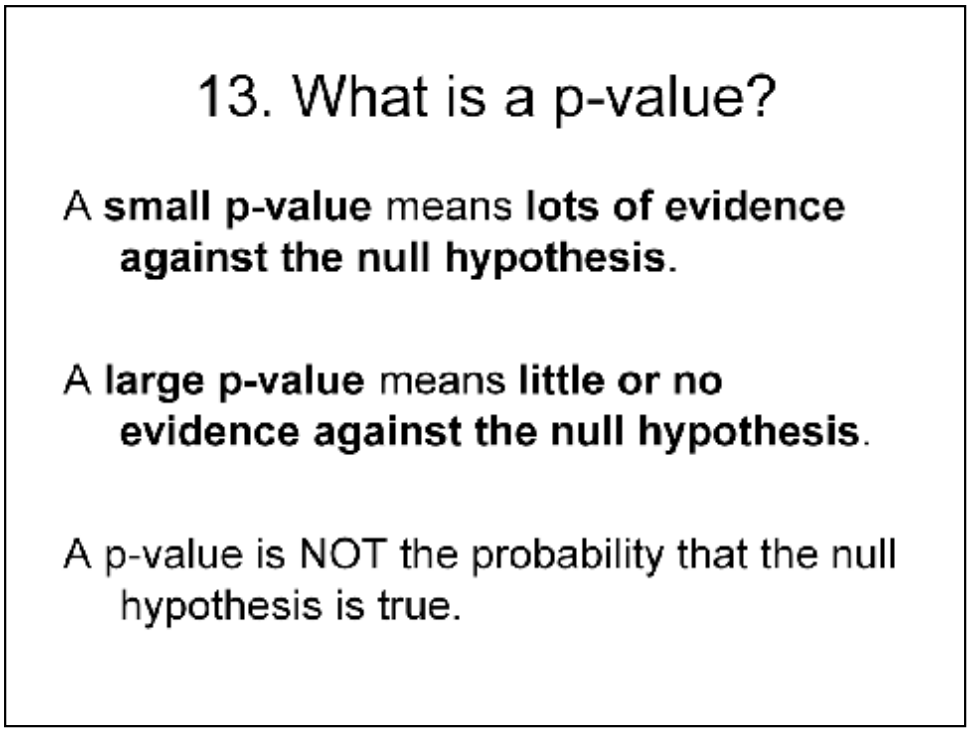
The p-value is also a measure of how likely we are to get a certain sample result or a result “more extreme,” assuming H0 is true. The type of hypothesis (right tailed, left tailed or two tailed) will determine what “more extreme” means.



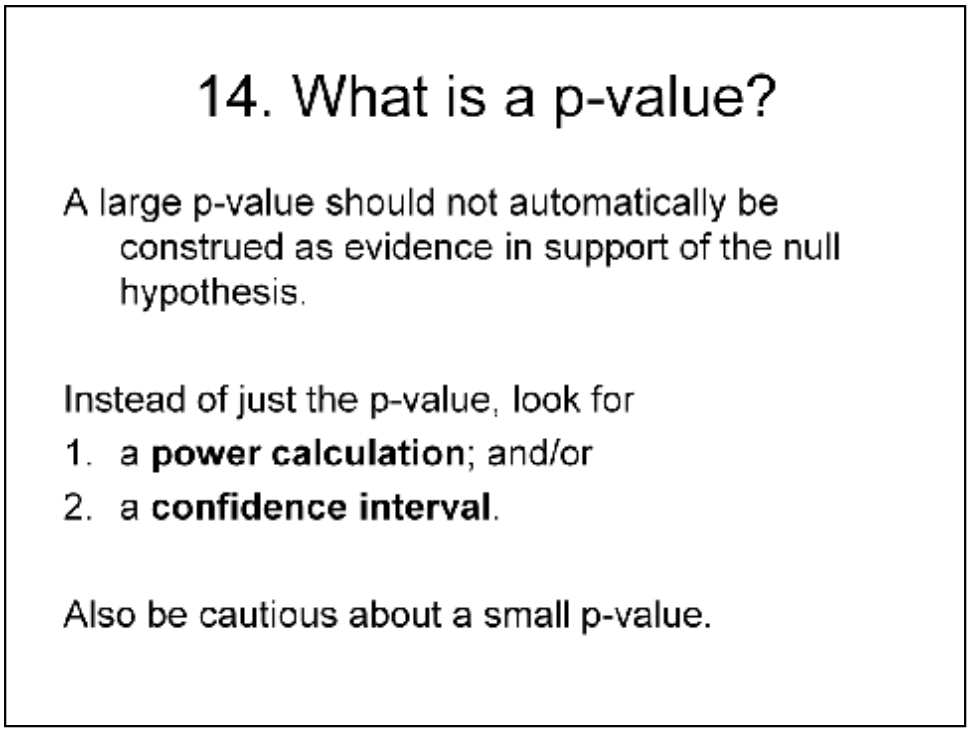
It is easiest to understand the p-value in a data set that is already at an extreme.

The data supports your belief because it is inconsistent with the assumption of a 50% adverse event rate. It would be like flipping a coin 12 times and getting heads each time.

The p-value, the probability of getting a sample result of **12 adverse events in 12 patients** assuming that the adverse event rate is 50%, is a measure of this inconsistency. **The p-value, 0.000244, is small enough that we would reject the hypothesis that the adverse event rate was only 50%.**



The p-value is frequently misinterpreted as the probability that the null hypothesis is false. The framework in which p-values are derived assume a fixed (non-random) pair of hypotheses. Randomness appears in the use of a random sample to characterize information about a fixed population. This approach to hypothesis testing considers any statement of the probability of the null hypothesis to be absurd. An alternative approach, Bayesian statistics, places prior probabilities on hypotheses and parameters.



Perhaps the failure to reject the null hypothesis was caused by an inadequate sample size. When you see a large p-value in a research study, you should also look for one of two things:

a **power calculation** that confirms that the sample size in that study was adequate for detecting a clinically relevant difference; and/or

a **confidence interval** that lies entirely within the range of clinical indifference.

You should also be cautious about a small p-value, but for different reasons. **In some situations, the sample size is so large that even differences that are trivial from a medical perspective can still achieve statistical significance**.

15. Practice exercise: interpret the p-values shown on the following pages.

Use the PICO format.

- P = Patient population
- I = Intervention
- C = Control/Comparison group
- O = Outcome

If you are having trouble reading the following abstracts, please go to

--> http://www.pmean.com/11/abstracts.html

It helps to orient yourself by finding the research hypothesis being tested. Not every research hypothesis will follow this PICO format, but it is a good starting point.

16. Winblad et al. **A six month double blind, randomized, placebo controlled study of a transdermal patch in Alzheimer's disease: rivastigmine patch versus capsule** International Journal of Geriatric Psychiatry. 2007;22(5):456-467. *Abstract: "Objectives: To compare the efficacy, safety and tolerability of a novel rivastigmine transdermal patch with conventional rivastigmine capsules and placebo in patients with Alzheimer's disease (AD). Methods: In this 24-week, multicenter, double-blind, double-dummy, placebo- and active-controlled trial, patients with probable AD were randomized to one of four treatment groups: 12 mg/day rivastigmine capsules; 10 cm2 (9.5 mg/24 h) rivastigmine patch; 20 cm2 (17.4 mg/24 h) rivastigmine patch; or placebo. Primary efficacy measures were the Alzheimer's Disease Assessment Scale--Cognitive subscale (ADAS-Cog) and Alzheimer's Disease Cooperative Study--Clinical Global Impression of Change (ADCS-CGIC). Results: One thousand one hundred and ninety five AD patients from 21 countries participated in the study. Treatment differences (vs placebo) on the ADAS-Cog at Week 24 in 10 cm2 patch, 20 cm2 patch and capsule groups were 1.6 (p = 0.005), 2.6 (p < 0.001) and 1.6 (p = 0.003). Treatment differences on the ADCS-CGIC were 0.3 (p = 0.01), 0.2 (p = 0.054) and 0.3 (p = 0.009). Comparison between the 10 cm2 patch and the capsule revealed non-inferiority. Rates of nausea in the 10 cm2 patch and capsule groups were 7.2% and 23.1%, respectively; rates of vomiting were 6.2% and 17.0%, respectively. Moderate or severe skin irritation occurred in <= 10% patients across the four patch sizes (5, 10, 15 and 20 cm2). Conclusions: The target dose of 10 cm2 rivastigmine patch provides efficacy similar to the highest doses of capsules with a superior tolerability profile. The transdermal patch with rivastigmine may offer convenience important to many caregivers and patients.* http://onlinelibrary.wiley.com/doi/10.1002/gps.1788/abstract.

P = patients with Alzheimer’s disease

I = 10 cm^2 or 20 cm^2 rivastigmine patch

C = patients taking 12 mg/day rivastigmine capsule and patients taking placebo

O = ADAS-Cog, ADCS-CGIC at 24 weeks. Nausea, vomitting.

If you are having trouble reading this, please go to

--> http://www.pmean.com/11/abstracts.html

17. Vernarelli et al. **Effect of Alzheimer disease genetic risk disclosure on dietary supplement use** Am. J. Clin. Nutr. 2010;91(5):1402-1407. *Abstract: "BACKGROUND: Genetic susceptibility testing for Alzheimer disease (AD) with APOE genotype disclosure is not recommended for clinical use but is available through direct-to-consumer (DTC) genetic testing companies. Little is known about whether APOE genotype disclosure would actually prompt changes in nutrition behaviors among at-risk individuals. OBJECTIVE: We studied the effect of APOE genotype disclosure for AD risk assessment on dietary supplement use in adults with a family history of AD. DESIGN: As part of a secondary analysis of data from the second Risk Evaluation and Education for Alzheimer's Disease Study, we examined the effect of genotype disclosure on health-behavior changes among 272 unaffected first-degree relatives of persons with AD. RESULTS: Overall, 16% of all participants reported a change in dietary supplement use after AD risk assessment. Participants who learned that they had at least one copy of the risk-increasing epsilon4 allele (epsilon4+) had 4.75 times the odds of reporting a change in dietary supplement use than did their counterparts who had an absence of the risk-increasing epsilon4 allele (epsilon4-) (95% CI: 2.23, 10.10; P < 0.0001) after adjustment for age, sex, race, baseline supplement use, randomization arm, and educational level. There were no significant differences between APOE epsilon4+ and epsilon4- participants in changes in overall diet, exercise, or medications. CONCLUSIONS: In this sample of first-degree relatives receiving genetic susceptibility testing for AD, an APOE epsilon4+ genotype status was positively associated with dietary supplement use after risk disclosure. Such changes occurred despite the absence of evidence that supplement use reduces the risk of AD. Given the expansion of DTC genetic tests, this study highlights the need for future studies in disease risk communication."* http://www.ncbi.nlm.nih.gov/pubmed/20219963.

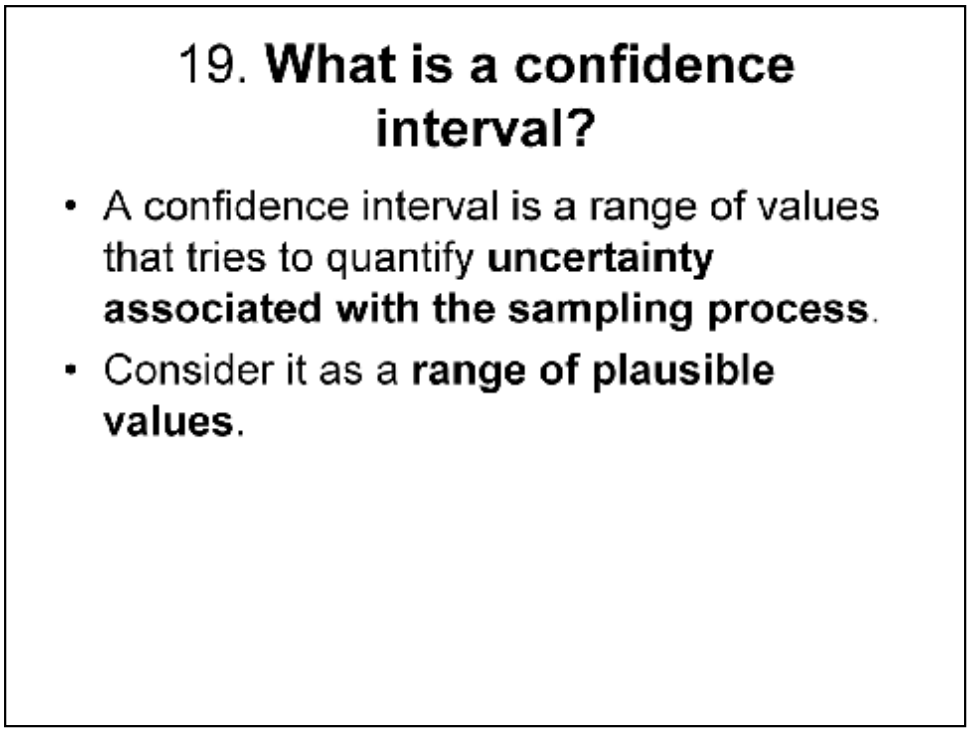
If you are having trouble reading this, please go to

--> http://www.pmean.com/11/abstracts.html

18. Rosenberg et al. **Sertraline for the treatment of depression in Alzheimer disease** Am J Geriatr Psychiatry. 2010;18(2):136-145. *Abstract: "OBJECTIVE: Depression is common in Alzheimer disease (AD), and antidepressants are commonly used for its treatment, however, evidence for antidepressant efficacy in this population is lacking. The authors conducted a multicenter, randomized, placebo-controlled trial titled "Depression in Alzheimer's Disease-2" to assess the efficacy and tolerability of sertraline for depression in AD. METHODS: One hundred thirty-one participants from five U.S. medical centers with mild-to-moderate AD (Mini-Mental State Examination scores 10-26) and depression of AD were randomized to double-blinded treatment with sertraline (N = 67) or placebo (N = 64), with a target dosage of 100 mg daily. Efficacy was assessed using logistic regressions and mixed effects models in an intention-to-treat analysis with imputation of missing data. Principal outcome measures were modified Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change (mADCS-CGIC), change in Cornell Scale for Depression in Dementia (CSDD) scores, and remission defined by both mADCS-CGIC score <= 2 and CSDD score <= 6. RESULTS: mADCS-CGIC ratings (odd ratio [OR = 1.01], 95% confidence interval [CI]: 0.52-1.97, p = 0.98), CSDD scores (median difference at 12 weeks 1.2, 95% CI: 1.65-4.05, p = 0.41), and remission at 12 weeks of follow-up (OR = 2.06, 95% CI: 0.84-5.04, p = 0.11) did not differ between sertraline (N = 67) and placebo (N = 64). Sertraline-treated patients experienced more adverse events, most notably gastrointestinal and respiratory, than placebo-treated patients. CONCLUSION: Sertraline did not demonstrate efficacy for the treatment depression symptoms in patients with AD. In addition, its use was associated with an increased incidence of adverse events. Thus, selective serotonin reuptake inhibitors may be of limited value for treating depression in patients with AD."* http://www.ncbi.nlm.nih.gov/pubmed/20087081.

If you are having trouble reading this, please go to

--> http://www.pmean.com/11/abstracts.html



We statisticians have a habit of **hedging our bets**. We always insert qualifiers into our reports, warn about all sorts of assumptions, and never admit to anything more extreme than probable. There's a famous saying: **"Statistics means never having to say you're certain**."

We qualify our statements, of course, because we are always **dealing with imperfect information**. In particular, we are often asked to make statements about a population (a large group of subjects) using information from a sample (a small, but carefully selected subset of this population). No matter how carefully this sample is selected to be a fair and unbiased representation of the population, **relying on information from a sample will always lead to some level of uncertainty**.

20. **What is a confidence interval?**

- Wide interval = poor precision
- Narrow interval = good precision
- It is unethical to conduct research if you know that your confidence interval will be so wide as to be uninformative.

A wide confidence interval implies poor precision; we can only specify plausible values to a broad and uninformative range.

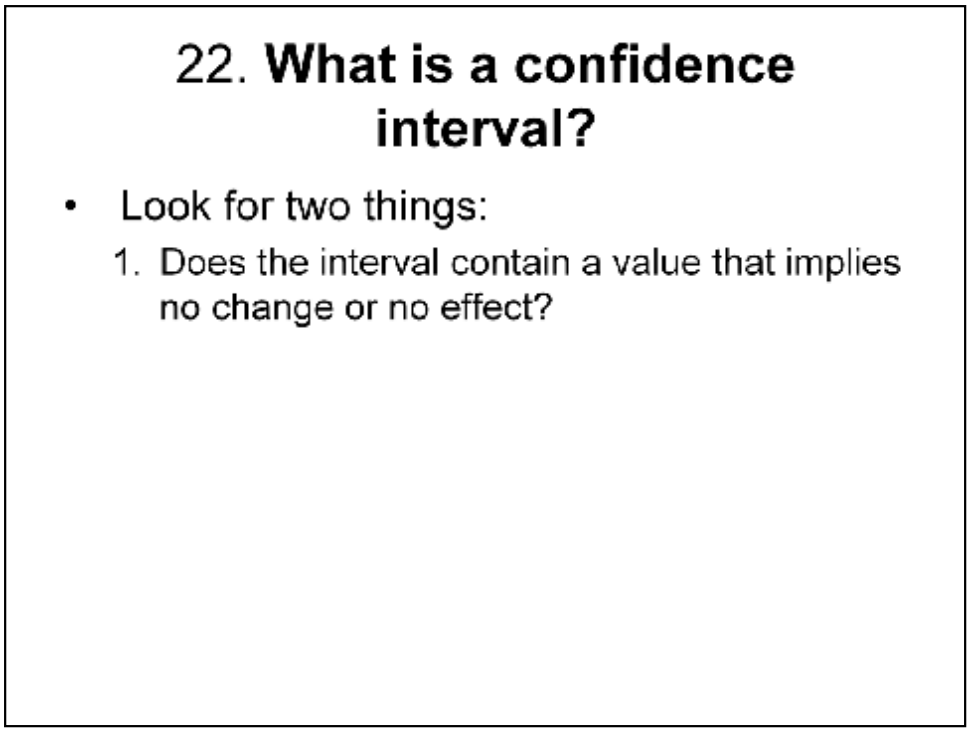
A narrow confidence interval implies good precision; we can place sharp limits on the range of plausible values.

21. **What is a confidence interval?**

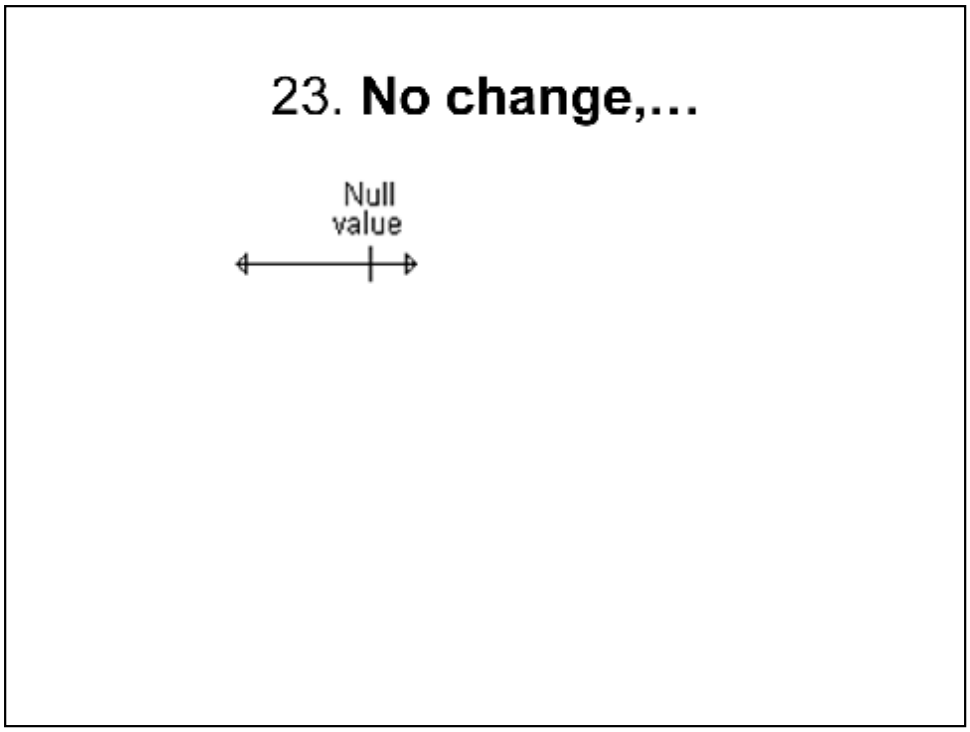
- Consider a recent study of homoeopathic treatment of pain and swelling after oral surgery.
  - P= patients undergoing oral surgery
  - I=homeopathic treatment
  - C=patients taking placebo.
  - O=swelling after 3 days.
- Homoeopathy led to 1 mm less swelling on average.
- The 95% confidence interval, however, ranged from -5.5 to 7.5 mm.

This interval implies that **neither a large improvement due to homoeopathy nor a large decrement could be ruled out**.

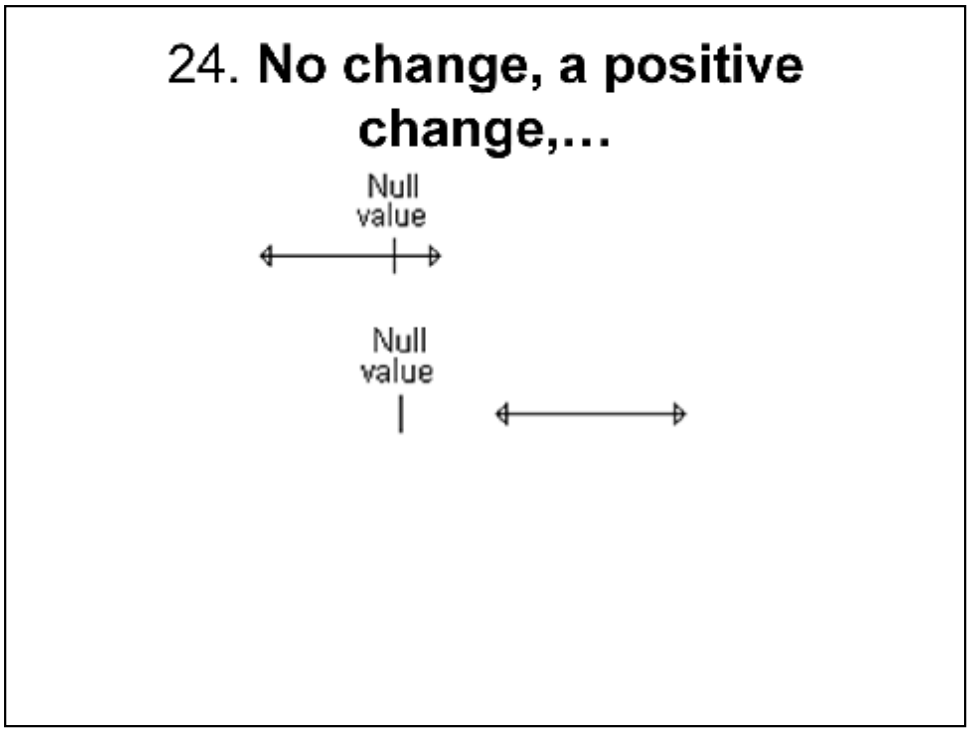
Reference is P. Lokken, P. A. Straumsheim, D. Tveiten, P. Skjelbred, C. F. Borchgrevink. Effect of homoeopathy on pain and other events after acute trauma: placebo controlled trial with bilateral oral surgery *BMJ*. 1995;310(6992):1439-1442. http://www.bmj.com/content/310/6992/1439.full



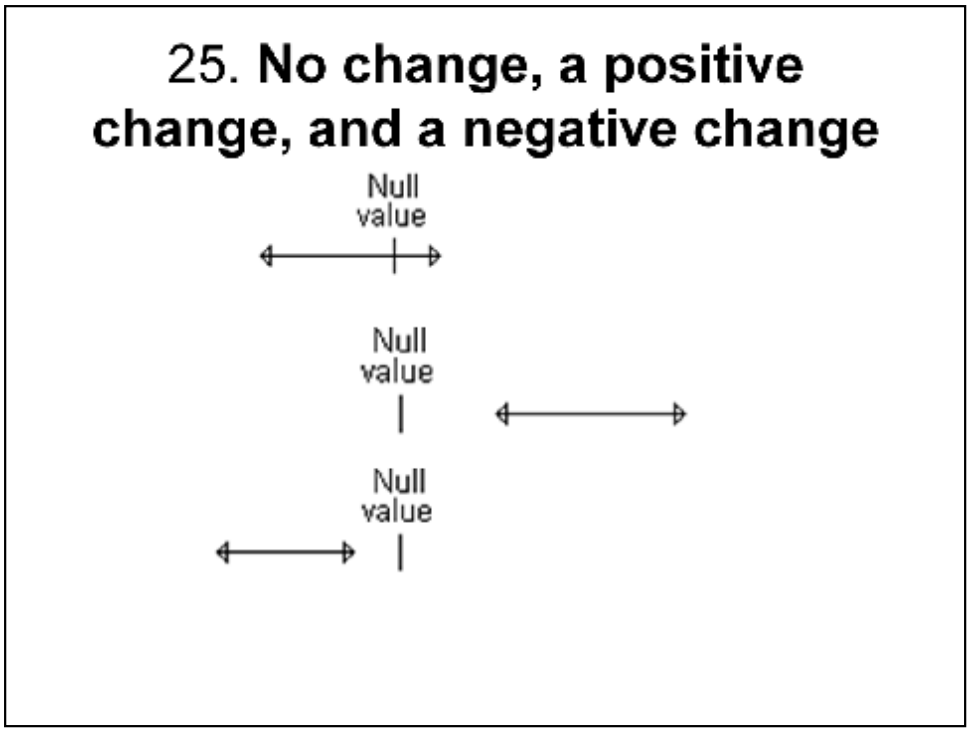
When you see a confidence interval in a published medical report, you should look for two things. First, **does the interval contain a value that implies no change or no effect**? For example, with a confidence interval for a difference look to see whether that interval includes zero. With a confidence interval for a ratio, look to see whether that interval contains one.



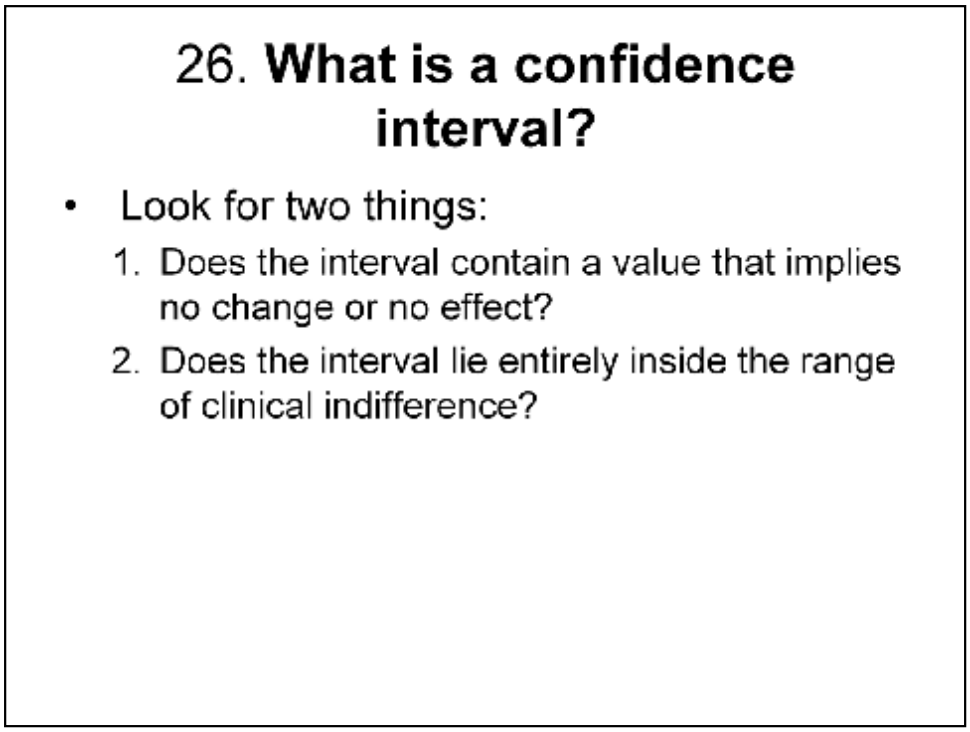
Here's an example of a confidence interval that contains the null value. **The interval shown below implies no statistically significant change**.



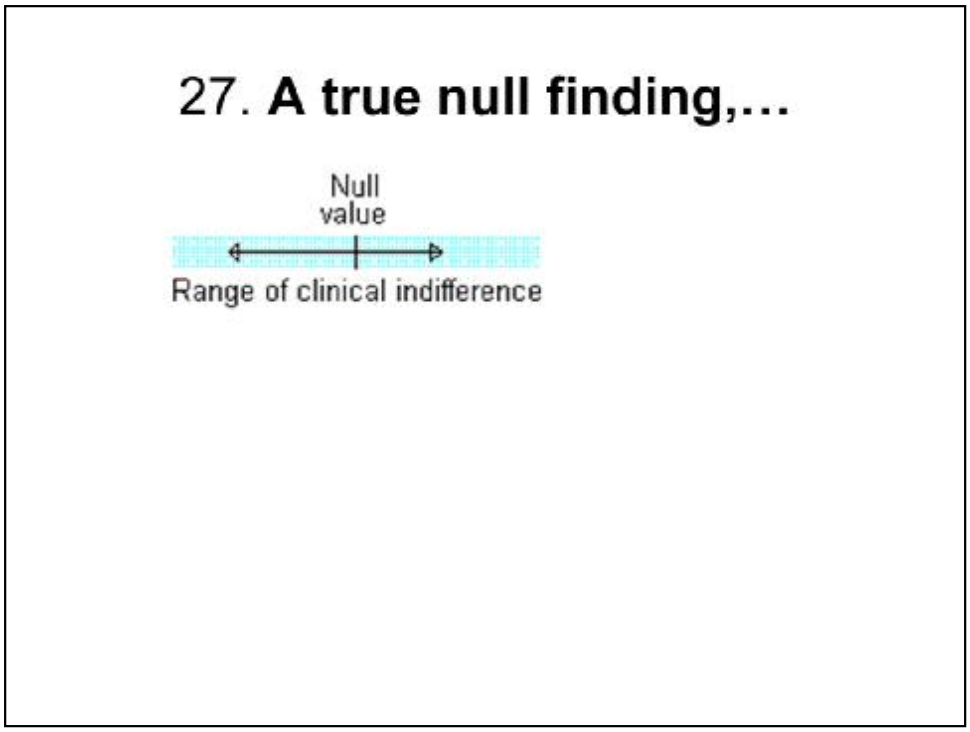
Here's an example of a confidence interval that excludes the null value. If we assume that larger implies better, then **the interval shown below would imply a statistically significant improvement**.



Here's a different example of a confidence interval that excludes the null value. **The interval shown below implies a statistically significant decline**.

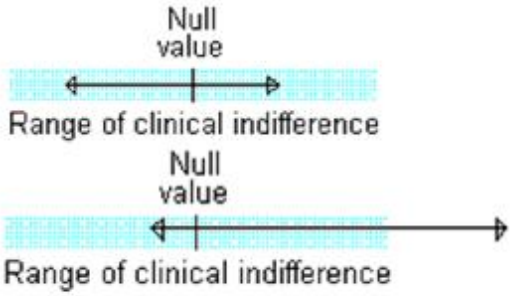


You should also see **whether the confidence interval lies partly or entirely within a range of clinical indifference**. Clinical indifference represents values of such a trivial size that you would not want to change your current practice.

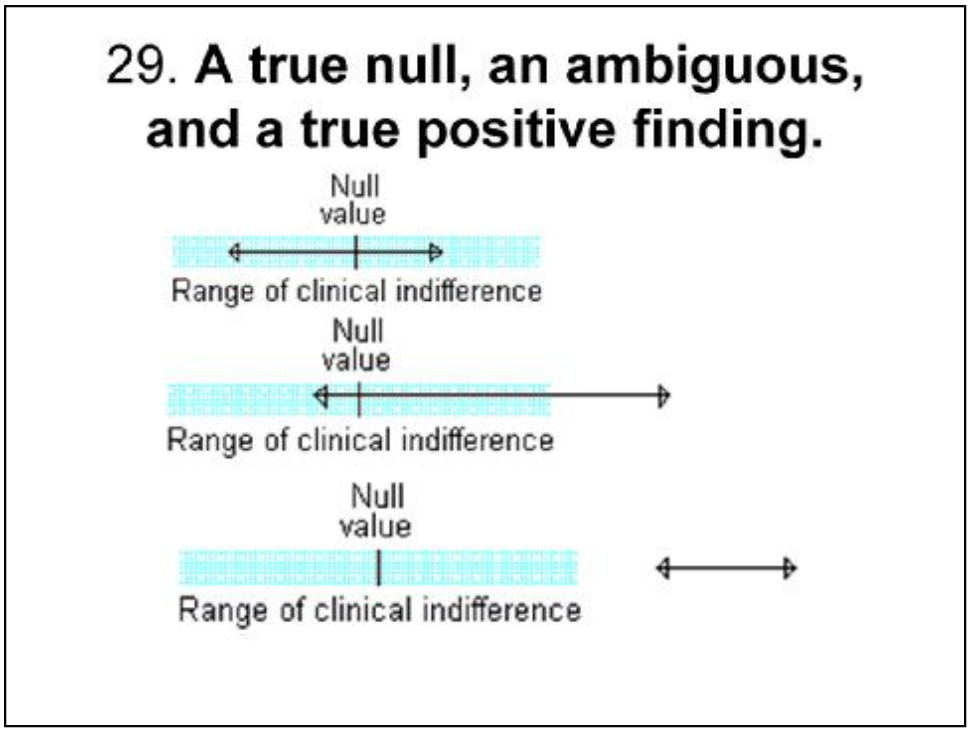


If a confidence interval is contained entirely within your range of clinical indifference, then you have clear and convincing evidence to **keep doing things the same way**.

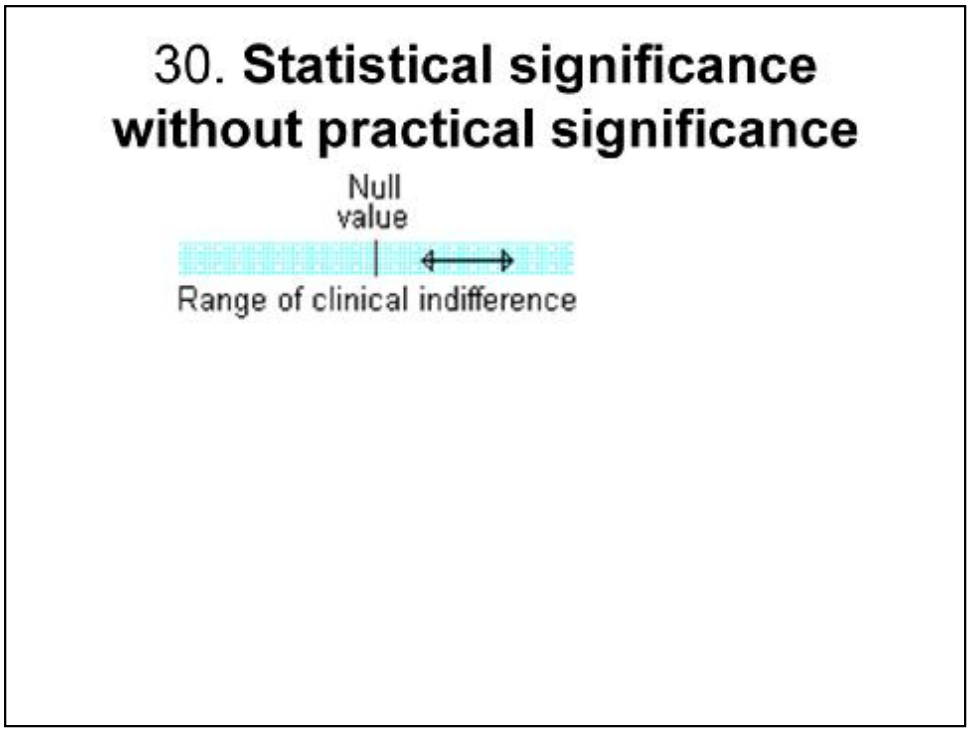
28. **A true null and an ambiguous finding, …**



One the other hand, **if part of the confidence interval lies outside the range of clinical indifference**, then you should consider the possibility that **the sample size is too small**.



Finally, if your **confidence interval excludes the null value and lies outside the range of clinical indifference**, then you have **both statistical and practical significance**.



If your **confidence interval excludes the null value but still lies entirely within the range of clinical indifference**, then you have a result with **statistical significance, but no practical significance**.

31. Practice exercise: interpret the confidence shown on the following pages.

If you are having trouble reading the following abstracts, please go to

--> http://www.pmean.com/11/abstracts.html

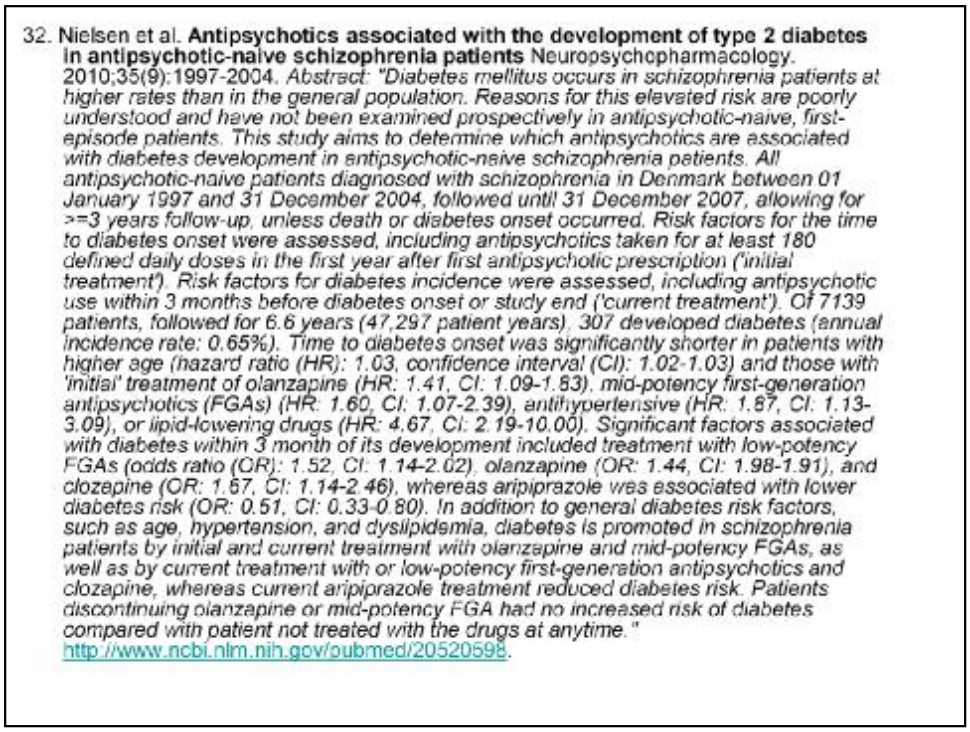
Recall that it helps to orient yourself by finding the research hypothesis being tested. Use the PICO format.

P = Patient population

I = Intervention

C = Control/Comparison group

O = Outcome



32. Nielsen et al. **Antipsychotics associated with the development of type 2 diabetes in antipsychotic-naive schizophrenia patients** Neuropsychopharmacology. 2010;35(9):1997-2004. *Abstract: "Diabetes mellitus occurs in schizophrenia patients at higher rates than in the general population. Reasons for this elevated risk are poorly understood and have not been examined prospectively in antipsychotic-naive, first-episode patients. This study aims to determine which antipsychotics are associated with diabetes development in antipsychotic-naive schizophrenia patients. All antipsychotic-naive patients diagnosed with schizophrenia in Denmark between 01 January 1997 and 31 December 2004, followed until 31 December 2007, allowing for >=3 years follow-up, unless death or diabetes onset occurred. Risk factors for the time to diabetes onset were assessed, including antipsychotics taken for at least 180 defined daily doses in the first year after first antipsychotic prescription ('initial treatment'). Risk factors for diabetes incidence were assessed, including antipsychotic use within 3 months before diabetes onset or study end ('current treatment'). Of 7139 patients, followed for 6.6 years (47,297 patient years), 307 developed diabetes (annual incidence rate: 0.65%). Time to diabetes onset was significantly shorter in patients with higher age (hazard ratio (HR): 1.03, confidence interval (CI): 1.02-1.03) and those with 'initial' treatment of olanzapine (HR: 1.41, CI: 1.09-1.83), mid-potency first-generation antipsychotics (FGAs) (HR: 1.60, CI: 1.07-2.39), antihypertensive (HR: 1.87, CI: 1.13-3.09), or lipid-lowering drugs (HR: 4.67, CI: 2.19-10.00). Significant factors associated with diabetes within 3 month of its development included treatment with low-potency FGAs (odds ratio (OR): 1.52, CI: 1.14-2.02), olanzapine (OR: 1.44, CI: 1.98-1.91), and clozapine (OR: 1.67, CI: 1.14-2.46), whereas aripiprazole was associated with lower diabetes risk (OR: 0.51, CI: 0.33-0.80). In addition to general diabetes risk factors, such as age, hypertension, and dyslipidemia, diabetes is promoted in schizophrenia patients by initial and current treatment with olanzapine and mid-potency FGAs, as well as by current treatment with or low-potency first-generation antipsychotics and clozapine, whereas current aripiprazole treatment reduced diabetes risk. Patients discontinuing olanzapine or mid-potency FGA had no increased risk of diabetes compared with patient not treated with the drugs at anytime."* http://www.ncbi.nlm.nih.gov/pubmed/20520598.

P = Anti-psychotic-naïve (first episode) schizophrenia patients

I = Multiple interventions: various antipsychotics, various demographics.

C = (unclear)

O = time to diabetes onset

If you are having trouble reading this, please go to

--> http://www.pmean.com/11/abstracts.html

33. Vernarelli et al. **Effect of Alzheimer disease genetic risk disclosure on dietary supplement use** Am. J. Clin. Nutr. 2010;91(5):1402-1407. *Abstract: "BACKGROUND: Genetic susceptibility testing for Alzheimer disease (AD) with APOE genotype disclosure is not recommended for clinical use but is available through direct-to-consumer (DTC) genetic testing companies. Little is known about whether APOE genotype disclosure would actually prompt changes in nutrition behaviors among at-risk individuals. OBJECTIVE: We studied the effect of APOE genotype disclosure for AD risk assessment on dietary supplement use in adults with a family history of AD. DESIGN: As part of a secondary analysis of data from the second Risk Evaluation and Education for Alzheimer's Disease Study, we examined the effect of genotype disclosure on health-behavior changes among 272 unaffected first-degree relatives of persons with AD. RESULTS: Overall, 16% of all participants reported a change in dietary supplement use after AD risk assessment. Participants who learned that they had at least one copy of the risk-increasing epsilon4 allele (epsilon4+) had 4.75 times the odds of reporting a change in dietary supplement use than did their counterparts who had an absence of the risk-increasing epsilon4 allele (epsilon4-) (95% CI: 2.23, 10.10; P < 0.0001) after adjustment for age, sex, race, baseline supplement use, randomization arm, and educational level. There were no significant differences between APOE epsilon4+ and epsilon4- participants in changes in overall diet, exercise, or medications. CONCLUSIONS: In this sample of first-degree relatives receiving genetic susceptibility testing for AD, an APOE epsilon4+ genotype status was positively associated with dietary supplement use after risk disclosure. Such changes occurred despite the absence of evidence that supplement use reduces the risk of AD. Given the expansion of DTC genetic tests, this study highlights the need for future studies in disease risk communication."* http://www.ncbi.nlm.nih.gov/pubmed/20219963.

If you are having trouble reading this, please go to

--> http://www.pmean.com/11/abstracts.html

34. Rosenberg et al. **Sertraline for the treatment of depression in Alzheimer disease** Am J Geriatr Psychiatry. 2010;18(2):136-145. *Abstract: "OBJECTIVE: Depression is common in Alzheimer disease (AD), and antidepressants are commonly used for its treatment, however, evidence for antidepressant efficacy in this population is lacking. The authors conducted a multicenter, randomized, placebo-controlled trial titled "Depression in Alzheimer's Disease-2" to assess the efficacy and tolerability of sertraline for depression in AD. METHODS: One hundred thirty-one participants from five U.S. medical centers with mild-to-moderate AD (Mini-Mental State Examination scores 10-26) and depression of AD were randomized to double-blinded treatment with sertraline (N = 67) or placebo (N = 64), with a target dosage of 100 mg daily. Efficacy was assessed using logistic regressions and mixed effects models in an intention-to-treat analysis with imputation of missing data. Principal outcome measures were modified Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change (mADCS-CGIC), change in Cornell Scale for Depression in Dementia (CSDD) scores, and remission defined by both mADCS-CGIC score <= 2 and CSDD score <= 6. RESULTS: mADCS-CGIC ratings (odd ratio [OR = 1.01], 95% confidence interval [CI]: 0.52-1.97, p = 0.98), CSDD scores (median difference at 12 weeks 1.2, 95% CI: 1.65-4.05, p = 0.41), and remission at 12 weeks of follow-up (OR = 2.06, 95% CI: 0.84-5.04, p = 0.11) did not differ between sertraline (N = 67) and placebo (N = 64). Sertraline-treated patients experienced more adverse events, most notably gastrointestinal and respiratory, than placebo-treated patients. CONCLUSION: Sertraline did not demonstrate efficacy for the treatment depression symptoms in patients with AD. In addition, its use was associated with an increased incidence of adverse events. Thus, selective serotonin reuptake inhibitors may be of limited value for treating depression in patients with AD."* http://www.ncbi.nlm.nih.gov/pubmed/20087081.

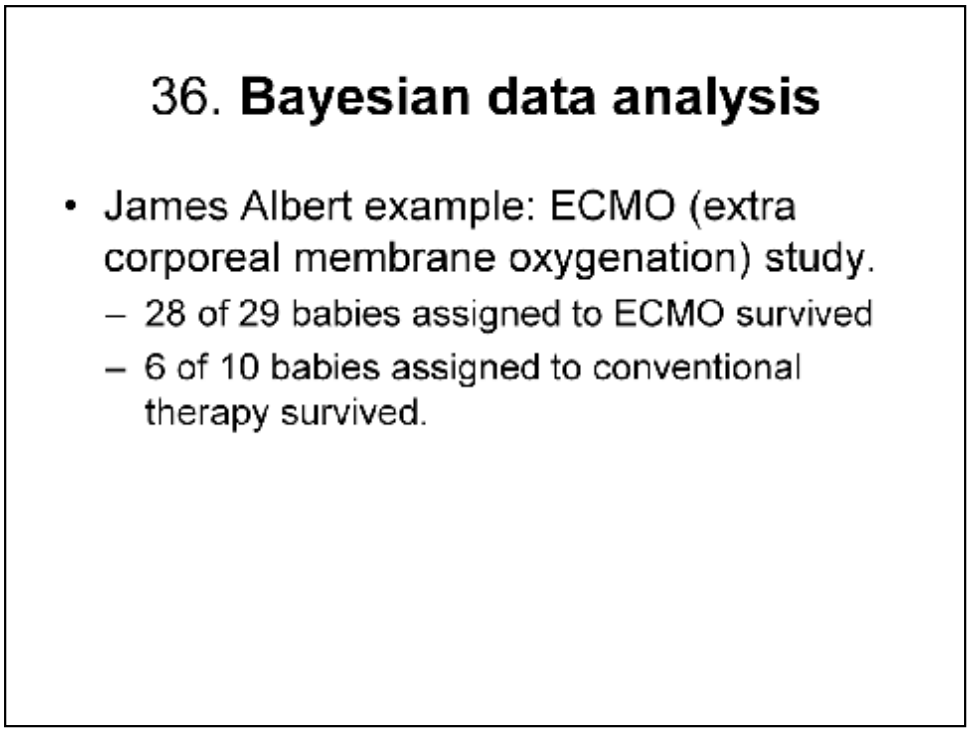
If you are having trouble reading this, please go to

--> http://www.pmean.com/11/abstracts.html

35. Weintraub et al. **Sertraline for the treatment of depression in Alzheimer disease: week-24 outcomes** Am J Geriatr Psychiatry. 2010;18(4):332-340. *Abstract: "BACKGROUND: Depression and antidepressant use are common in Alzheimer disease (AD), but the effect of antidepressant treatment for depression on longer term outcomes is unknown. The authors report the Week-24 outcomes of patients who participated in a 12-week efficacy study of sertraline for depression of AD. METHODS: One hundred thirty-one participants (sertraline = 67, placebo = 64) with mild-moderate AD and depression participated in the study. Patients who showed improvement on the modified Alzheimer's Disease Cooperative Study Clinical Global Impression-Change (mADCS-CGIC) after 12 weeks of randomized treatment with sertraline or placebo continued double-blinded treatment for an additional 12 weeks. Depression response and remission at 24 weeks were based on mADCS-CGIC score and change in Cornell Scale for Depression in Dementia (CSDD) score. Secondary outcome measures included time to remission, nonmood neuropsychiatric symptoms, global cognition, function, and quality of life. RESULTS: One hundred seventeen (89.3%) participants completed all study assessments and 74 (56.5%; sertraline = 38, placebo = 36) completed all 24 weeks on randomized treatment. By 24 weeks, there were no between-group differences in depression response (sertraline = 44.8%, placebo = 35.9%; odds ratio [95% CI] = 1.23 [0.64-2.35]), change in CSDD score (median difference = 0.6 [95% CI: -2.26 to 3.46], chi2 [df = 2] = 1.03), remission rates (sertraline = 32.8%, placebo = 21.8%; odds ratio [95% CI] = 1.61 [0.70-3.68]), or secondary outcomes. Common selective serotonin reuptake inhibitor-associated adverse events, specifically diarrhea, dizziness, and dry mouth, and pulmonary serious adverse events were more frequent in sertraline-randomized patients than in placebo subjects. CONCLUSIONS: Sertraline treatment is not associated with delayed improvement between 12 and 24 weeks of treatment and may not be indicated for the treatment of depression of AD."* http://www.ncbi.nlm.nih.gov/pubmed/20220589.

If you are having trouble reading this, please go to

--> http://www.pmean.com/11/abstracts.html



There's a wonderful example of Bayesian data analysis at work that is simple and fun. It's taken directly from an article by Jim Albert in the Journal of Statistics Education (1995, vol. 3 no. 3) which is available on the web at

www.amstat.org/publications/jse/v3n3/albert.html

This data is from the second example in this paper.

37. **Bayesian data analysis**

Classic formula for Bayesian data analysis:

$$P(H|E) = P(E|H)\,P(H) / P(E)$$

- H = hypothesis
- E = evidence (data)
- P, of course, stands for probability.

This is the classic formulation used for Bayesian data analysis. You can find a good description on Wikipedia.

38. **Bayesian data analysis**

- P(H) is the prior probability.
- The prior probability represents the degree of belief that you have in a particular hypothesis prior to collection of your data.

The first step in any Bayesian data analysis is specifying P(H) which is the prior probability. Specifying the prior probability is probably the one aspect of Bayesian data analysis that causes the most controversy.

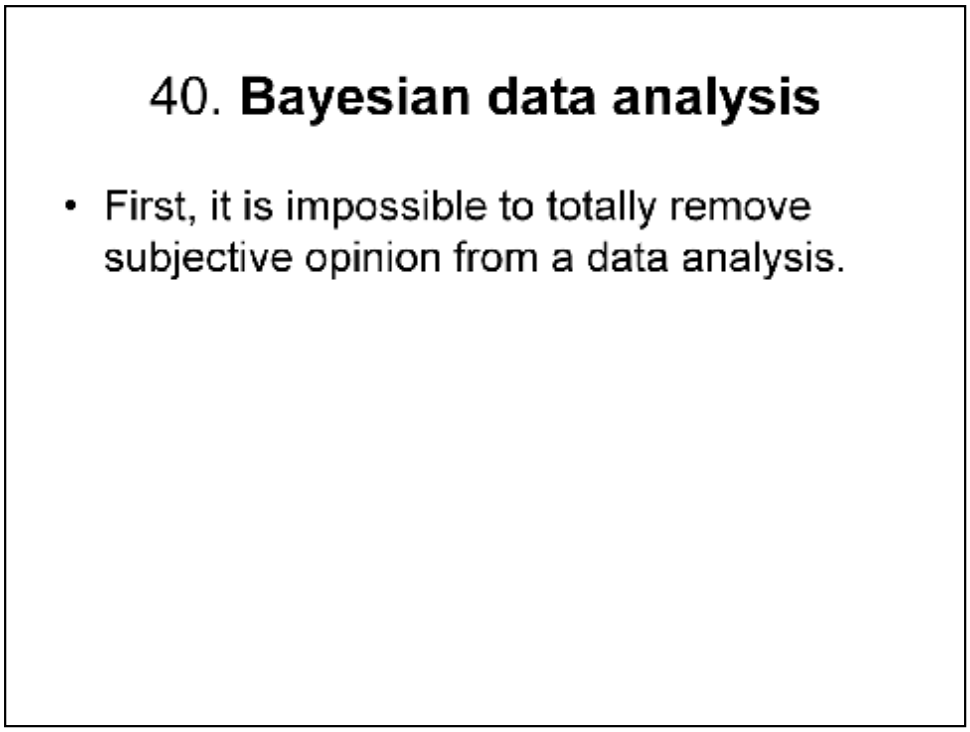
39. **Bayesian data analysis**

- The prior distribution can incorporate data from previous related studies or it can incorporate subjective impressions of the researcher.

The prior distribution can incorporate knowledge stemming from previous studies. No study is ever done in isolation, and savvy readers will always place a research articles findings in the context of what has already been shown. This is done in an informal and subjective fashion. Bayesian data analysis allows you to formalize this approach if you like.

The prior distribution can also incorporate subjective opinions of the researcher. What!?! you're saying right now. Aren't statistics supposed to remove the need for subjective opinions?

Actually, a bit of subjectivity is a good thing.

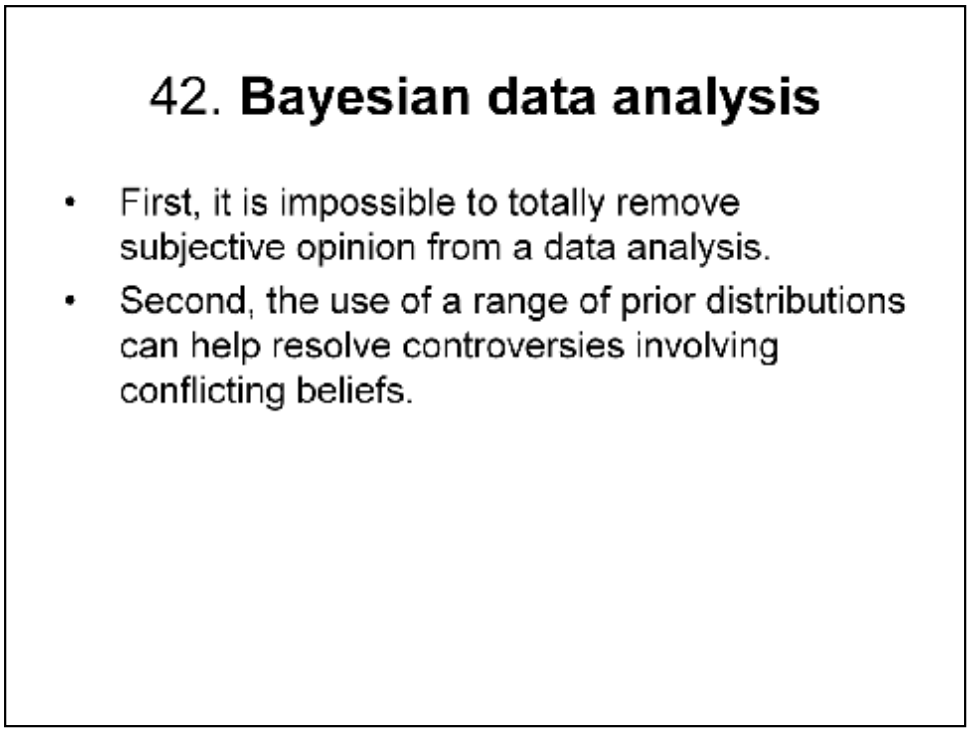


You can't do research without adopting some informal rules. These rules may be reasonable, they may be supported to some extent by empirical data, but they are still applied in a largely subjective fashion.

41. **Bayesian data analysis**

Here are some of the subjective beliefs that I use in my work:

1. you should always prefer a simple model to a complex model if both predict the data with the same level of precision.
2. you should be cautious about any subgroup finding that was not pre-specified in the research protocol.
3. if you can find a plausible biological mechanism, that adds credibility to your results.

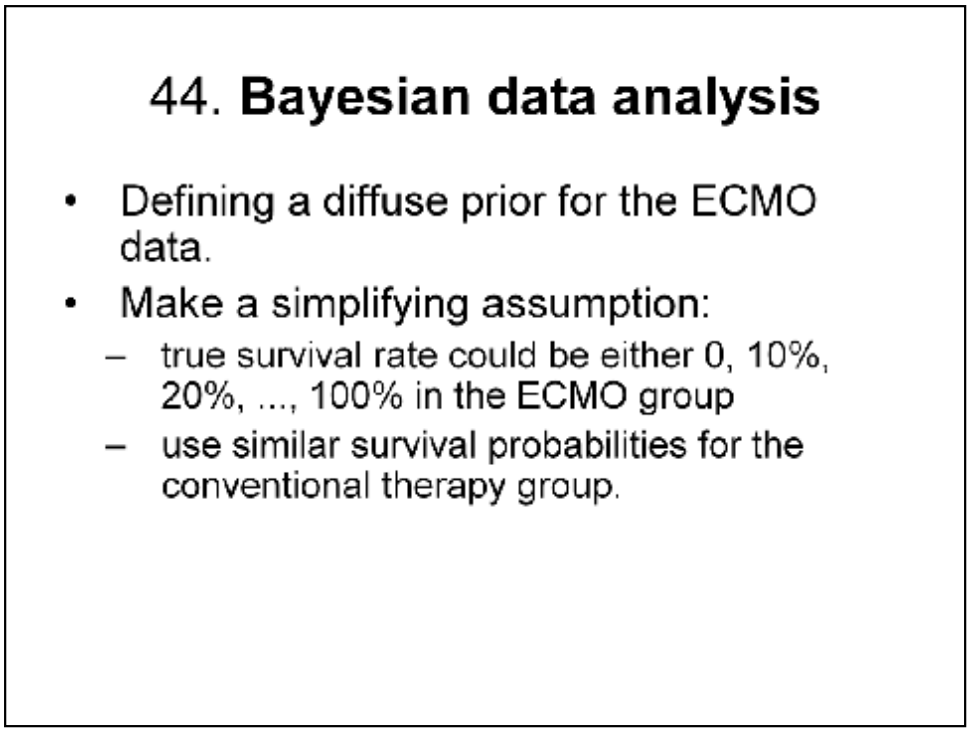


For example, an important research question is whether a research finding should "close the book" to further research. If data indicates a negative result, and this result is negative even using an optimistic prior probability, then all researchers, even those with the most optimistic hopes for the therapy, should move on.

43. **Bayesian data analysis**

- First, it is impossible to totally remove subjective opinion from a data analysis.
- Second, the use of a range of prior distributions can help resolve controversies involving conflicting beliefs.
- Third, while Bayesian data analysis allows you to incorporate subjective opinions into your prior probability, it does not require you to incorporate subjectivity.

Many Bayesian data analyses use what it called a diffuse or non-informative prior distribution. This is a prior distribution that is neither optimistic nor pessimistic, but spreads the probability more or less evenly across all hypotheses.



This is not an optimal assumption, but it isn't terrible either, and it allows us to see some of the calculations in action.

With 11 probabilities for ECMO and 11 probabilities for conventional therapy, we have 121 possible combinations. How should we arrange those probabilities? One possibility is to assign half of the total probability to combinations where the probabilities are the same for ECMO and conventional therapy and the remaining half to combinations where the probabilities are different. Split each of these probabilities evenly over all the combinations.

45. P (H|E) = P(E|H) **P(H)** / P(E)

- This is P(H). For simplicity, we multiplied each probability by 1000 and rounded.

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	45	5	5	5	5	5	5	5	5	5	5
0.1	5	45	5	5	5	5	5	5	5	5	5
0.2	5	5	45	5	5	5	5	5	5	5	5
0.3	5	5	5	45	5	5	5	5	5	5	5
0.4	5	5	5	5	45	5	5	5	5	5	5
0.5	5	5	5	5	5	45	5	5	5	5	5
0.6	5	5	5	5	5	5	45	5	5	5	5
0.7	5	5	5	5	5	5	5	45	5	5	5
0.8	5	5	5	5	5	5	5	5	45	5	5
0.9	5	5	5	5	5	5	5	5	5	45	5
1	5	5	5	5	5	5	5	5	5	5	45

There are 11 cells along the diagonal. 0.5 / 11 = 0.0045.

There are 110 cells off the diagonal. 0.5 / 110 = 0.0005.

46. P (H|E) = **P(E|H)** P(H) / P(E)

- The second step in a Bayesian data analysis is to calculate P(E | H).

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	0	0	0	0	0	0	0	0	0	0	0
0.1	0	0	0	0	0	0	0	0	0	0	0
0.2	0	0	0	0	0	0	0	0	0	1	0
0.3	0	0	0	0	0	0	0	0	0	6	0
0.4	0	0	0	0	0	0	0	0	1	17	0
0.5	0	0	0	0	0	0	0	0	2	31	0
0.6	0	0	0	0	0	0	0	0	3	38	0
0.7	0	0	0	0	0	0	0	0	2	30	0
0.8	0	0	0	0	0	0	0	0	1	13	0
0.9	0	0	0	0	0	0	0	0	0	2	0
1	0	0	0	0	0	0	0	0	0	0	0

P(E | H) is the probability of the observed data under each hypothesis. There are formulas for this, using the binomial distribution. I used Excel to calculate these probabilities for me. Here's an example of the formula I used.

```
=binomdist(28,29,0.9,FALSE)*binomdist(
6,10,0.6,FALSE)
```

47. P (H|E) = **P(E|H) P(H)** / P(E)

- Now multiply the prior probability by the likelihood.

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	0	0	0	0	0	0	0	0	0	0	0
0.1	0	0	0	0	0	0	0	0	0	0	0
0.2	0	0	0	0	0	0	0	0	0	4	0
0.3	0	0	0	0	0	0	0	0	2	25	0
0.4	0	0	0	0	0	0	0	0	6	77	0
0.5	0	0	0	0	0	0	0	0	10	141	0
0.6	0	0	0	0	0	0	0	1	13	173	0
0.7	0	0	0	0	0	0	0	4	10	138	0
0.8	0	0	0	0	0	0	0	0	45	67	0
0.9	0	0	0	0	0	0	0	0	1	77	0
1	0	0	0	0	0	0	0	0	0	0	0

By multiplying the probabilities together, you combine information from your prior belief and information from the data. The largest values will be from combinations of ECMO and conventional survival probabilities that have a reasonably large prior probability.

48. P (H|E) = P(E|H) P(H) / P(E)

- These numbers do not add up to 1, so we need to rescale them.

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	0	0	0	0	0	0	0	0	0	0	0
0.1	0	0	0	0	0	0	0	0	0	0	0
0.2	0	0	0	0	0	0	0	0	0	5	0
0.3	0	0	0	0	0	0	0	0	2	32	0
0.4	0	0	0	0	0	0	0	0	7	97	0
0.5	0	0	0	0	0	0	0	0	13	178	0
0.6	0	0	0	0	0	0	0	1	16	218	0
0.7	0	0	0	0	0	0	0	5	13	174	0
0.8	0	0	0	0	0	0	0	0	57	84	0
0.9	0	0	0	0	0	0	0	0	1	97	0
1	0	0	0	0	0	0	0	0	0	0	0

49. **Bayesian data analysis**

- P(Null | Data) = (0+0+...+5+57+97+0) = 159 / 1000.
- This is substantial decline from prior on the null (500 / 1000).
- The data has greatly (but not completely) disuaded us from our belief in the null hypothesis.

How likely are we to believe the hypothesis that ECMO and conventional therapy have the same survival rates? Just add the cells along the diagonal (0+0+...+5+57+97+0) to get 159 out of a thousand. Prior to collecting the data, we placed the probability that the two rates were equal at 500 out of a thousand, so the data has greatly (but not completely) dissuaded us from this belief.

50. **Bayesian data analysis**

You can calculate the probability that

- ECMO is exactly 10% better than conventional therapy
  - 0+0+...+1+13+84+0 = 98 /1000,
- ECMO is exactly 20% better
  - 0+0+...+13+218+0 = 231 / 1000,
- exactly 30% better
  - 0+0+...+7+178+0 = 185 /1000,
- and so forth.

51. Bayesian data analysis

- Here's something fun that Dr. Albert didn't show. You could take each of the cells in the table, compute a ratio of survival rates and then calculate the median of these ratios as 1.5.

Relative risk	Prob.	Cumul.	
0.8 / 0.9 = 0.89	1	1	
0.9 / 0.9 = 1.00	97	98	
0.8 / 0.8 = 1.00	57	155	
0.7 / 0.7 = 1.00	5	160	
0.9 / 0.8 = 1.12	84	244	
0.8 / 0.7 = 1.14	13	257	
0.7 / 0.6 = 1.17	1	258	
0.9 / 0.7 = 1.29	174	432	
0.8 / 0.6 = 1.33	16	448	
0.9 / 0.6 = 1.50	218	666	<< Median
0.8 / 0.5 = 1.60	13	679	
0.9 / 0.5 = 1.80	178	857	
0.8 / 0.4 = 2.00	7	864	
0.9 / 0.4 = 2.25	97	961	
0.8 / 0.3 = 2.67	2	963	
0.9 / 0.3 = 3.00	32	995	
0.9 / 0.2 = 4.50	5	1000	

You can get assign probability statements to the ratio of the survival rates as well. There are 840 chances out of a thousand (1000 – 160) that the ratio is greater than 1.0, but only 136 chances out of a thousand (1000 – 864) that the ratio is greater than 2.0.

52. **Bayesian data analysis**

- Here's a prior distribution that utilizes historic information.

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	Total
0	0	0	0	0	0	0	0	0	0	0	0	0
0.1	0	50	50	48	39	34	25	17	11	6	0	280
0.2	0	54	54	51	42	36	27	18	12	6	0	300
0.3	0	38	38	36	29	25	19	13	8	4	0	210
0.4	0	22	22	20	17	14	11	7	5	2	0	120
0.5	0	11	11	10	8	7	5	4	2	1	0	60
0.6	0	5	5	5	4	4	3	2	1	1	0	30
0.7	0	0	0	0	0	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0	0	0	0	0	0
0.9	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0
Total	0	180	180	170	140	120	90	60	40	20	0	

Dr. Albert goes on to show an informative prior distribution. There is a fair amount of data to indicate that the survival rate for the conventional therapy is somewhere between 10% and 30%, but little or no data about the survival rates under ECMO.

53. Repeat of pop quiz

A research paper computes a p-value of 0.45. How would you interpret this p-value?

1. Strong evidence for the null hypothesis
2. Strong evidence for the alternative hypothesis
3. Little or no evidence for the null hypothesis
4. Little or no evidence for the alternative hypothesis
5. More than one answer above is correct.
6. I do not know the answer.

54. Repeat of pop quiz

A research paper computes a confidence interval for a relative risk of 0.82 to 3.94. What does this confidence interval tell you.

1. The result is statistically significant and clinically important.
2. The result is not statistically significant, but is clinically important.
3. The result is statistically significant, but not clinically important.
4. The result is not statistically significant, and not clinically important.
5. The result is ambiguous.
6. I do not know the answer.

55. Repeat of pop quiz

A Bayesian data analysis can incorporate subjective opinions through the use of

1. Bayes rule.
2. data shrinkage.
3. a prior distribution.
4. a posterior distribution.
5. p-values.
6. I do not know the answer.