Statistical Inference with R

Continuous Variables

•••

GW Libraries Workshop

Dan Kerchner ~ September 14, 2023

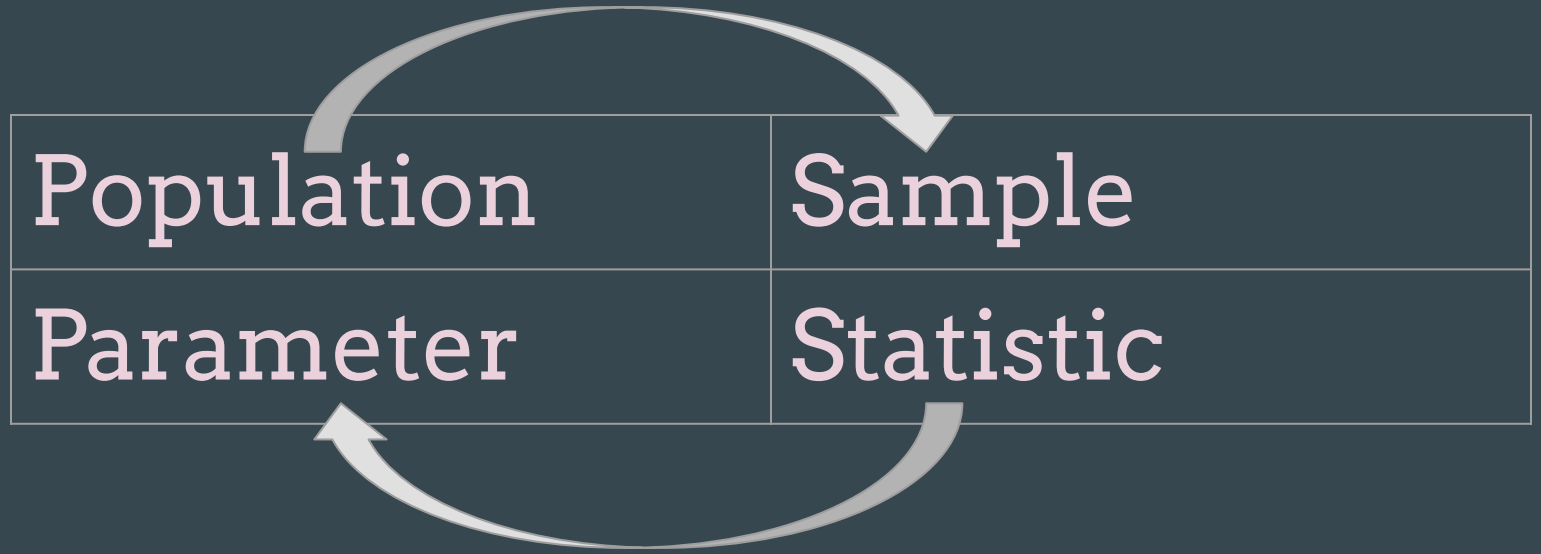
go.gwu.edu/rstats

Logistics

- Just speak up OR use the Zoom chat
- Plan for 1 brief  break

Super-Brief Review of Inference for Numerical (Continuous) Variables

High-Level Objective



Two forms of inference

Confidence Interval

95% CI for y = (2.3, 3.1)

Hypothesis Testing

H_0: $y = y_0$ ← Null Hypothesis
H_A: $y \neq y_0$ (or $y < y_0$ or $y > y_0$) ← Alternative Hypothesis

p-value: Chance that we are rejecting H_0 when we should not
α ← significance level: Tolerance for rejecting H_0 when we should not

Quick refresher on t tests and ANOVA

1-sample t-test:
What is the chance we would see this distribution if the true mean was μ?

2-sample t-test:
What is the chance we would see this data if both groups are actually drawn from the same distribution?

ANOVA:
What is the chance we would see this data if all "n" groups are generated by the same distribution?

Quick refresher on t tests and ANOVA

Two types of t tests:

- Paired: 2 observations (for example, before and after) of the same subject (with no group).
- Unpaired: single observation of the same group. Two types of tests:
  - Assume equal variances
  - Do not assume equal variances (more conservative assumption)

More on Types of t tests

- Two-tailed t test (more conservative)
  - Is (group A mean > group B mean or group A mean < group B mean)
- One-tailed t test (less conservative)
  - Is group A mean > group B

    or
  - Is group A mean < group B

Prerequisites ~ Assumptions for Validity

For t test and ANOVA:

- observations are independent within each group & across groups
- $n \geq 30$ (for t test)
- data are normally distributed

For 2-sample t test:

- also need ~ equal variances (otherwise ok, use an approximation)

When the assumptions are not satisfied, we may use other approaches (nonparametric tests, bootstrapping, etc.)

Goals

Today's Goal

- Learn to use R to read in data and conduct hypothesis tests for **continuous** measures
  - Checking assumptions
  - Visualizing
  - Computing p-values and confidence intervals

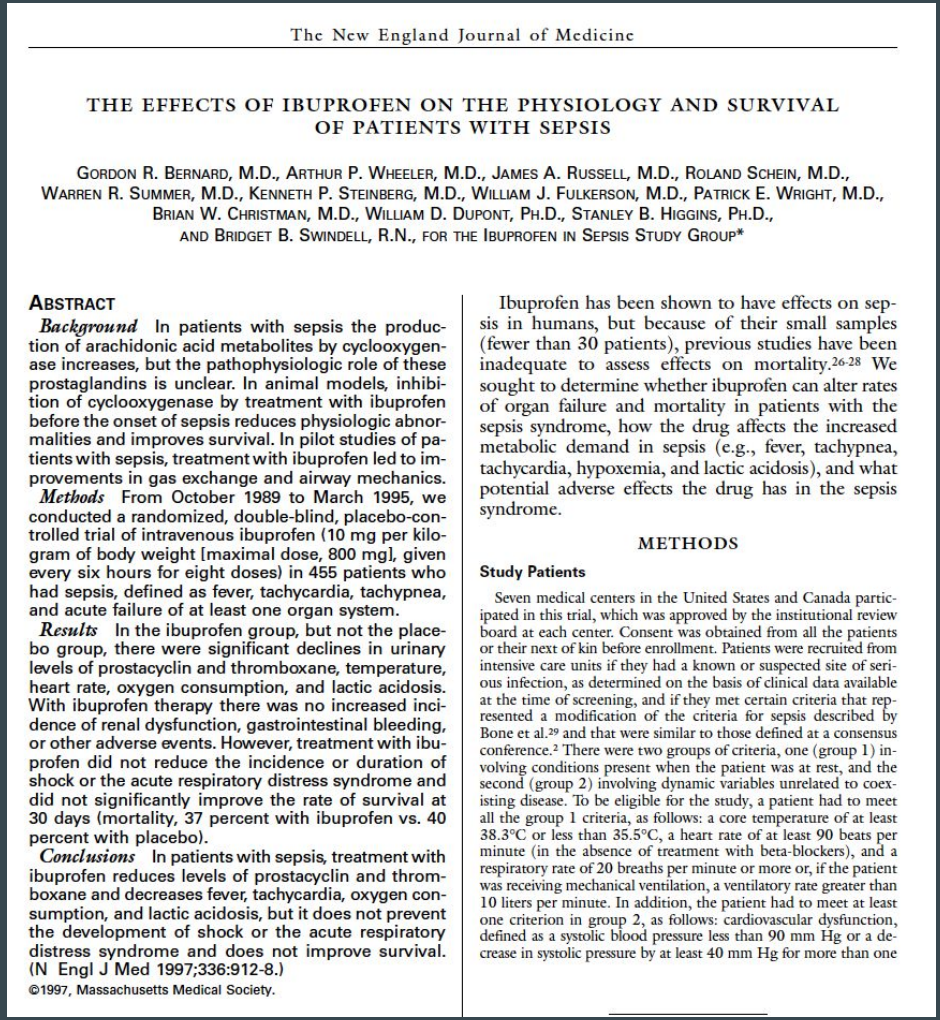
Today: 4 Scenarios

- Single-sample T Test
- Paired T Test
- 2-Sample T Test
- ANOVA (inference for means from more than 2 independent groups)

Today's Data Set

Bernard, G.R., Wheeler, A.P., Russell, J.A., Schein, R., Summer, W.R., Steinberg, K.P., et al. The effects of ibuprofen on the physiology and survival of patients with sepsis. The Ibuprofen in Sepsis Study Group. N. Engl. J. Med. 1997, 336: 912-8.

doi.org/10.1056/NEJM199703273361303



The New England Journal of Medicine

THE EFFECTS OF IBUPROFEN ON THE PHYSIOLOGY AND SURVIVAL OF PATIENTS WITH SEPSIS

GORDON R. BERNARD, M.D., ARTHUR P. WHEELER, M.D., JAMES A. RUSSELL, M.D., ROLAND SCHEIN, M.D., WARREN R. SUMMER, M.D., KENNETH P. STEINBERG, M.D., WILLIAM J. FULKERSON, M.D., PATRICK E. WRIGHT, M.D., BRIAN W. CHRISTMAN, M.D., WILLIAM D. DUPONT, PH.D., STANLEY B. HIGGINS, PH.D., AND BRIDGET B. SWINDELL, R.N., FOR THE IBUPROFEN IN SEPSIS STUDY GROUP*

ABSTRACT

Background In patients with sepsis the production of arachidonic acid metabolites by cyclooxygenase increases, but the pathophysiologic role of these prostaglandins is unclear. In animal models, inhibition of cyclooxygenase by treatment with ibuprofen before the onset of sepsis reduces physiologic abnormalities and improves survival. In pilot studies of patients with sepsis, treatment with ibuprofen led to improvements in gas exchange and airway mechanics.

Methods From October 1989 to March 1995, we conducted a randomized, double-blind, placebo-controlled trial of intravenous ibuprofen (10 mg per kilogram of body weight [maximal dose, 800 mg], given every six hours for eight doses) in 455 patients who had sepsis, defined as fever, tachycardia, tachypnea, and acute failure of at least one organ system.

Results In the ibuprofen group, but not the placebo group, there were significant declines in urinary levels of prostacyclin and thromboxane, temperature, heart rate, oxygen consumption, and lactic acidosis. With ibuprofen therapy there was no increased incidence of renal dysfunction, gastrointestinal bleeding, or other adverse events. However, treatment with ibuprofen did not reduce the incidence or duration of shock or the acute respiratory distress syndrome and did not significantly improve the rate of survival at 30 days (mortality, 37 percent with ibuprofen vs. 40 percent with placebo).

Conclusions In patients with sepsis, treatment with ibuprofen reduces levels of prostacyclin and thromboxane and decreases fever, tachycardia, oxygen consumption, and lactic acidosis, but it does not prevent the development of shock or the acute respiratory distress syndrome and does not improve survival. (N Engl J Med 1997;336:912-8.)

Ibuprofen has been shown to have effects on sepsis in humans, but because of their small samples (fewer than 30 patients), previous studies have been inadequate to assess effects on mortality.[26-28] We sought to determine whether ibuprofen can alter rates of organ failure and mortality in patients with the sepsis syndrome, how the drug affects the increased metabolic demand in sepsis (e.g., fever, tachypnea, tachycardia, hypoxemia, and lactic acidosis), and what potential adverse effects the drug has in the sepsis syndrome.

METHODS

Study Patients

Seven medical centers in the United States and Canada participated in this trial, which was approved by the institutional review board at each center. Consent was obtained from all the patients or their next of kin before enrollment. Patients were recruited from intensive care units if they had a known or suspected site of serious infection, as determined on the basis of clinical data available at the time of screening, and if they met certain criteria that represented a modification of the criteria for sepsis described by Bone et al.[29] and that were similar to those defined at a consensus conference.[2] There were two groups of criteria, one (group 1) involving conditions present when the patient was at rest, and the second (group 2) involving dynamic variables unrelated to coexisting disease. To be eligible for the study, a patient had to meet all the group 1 criteria, as follows: a core temperature of at least 38.3°C or less than 35.5°C, a heart rate of at least 90 beats per minute (in the absence of treatment with beta-blockers), and a respiratory rate of 20 breaths per minute or more or, if the patient was receiving mechanical ventilation, a ventilatory rate greater than 10 liters per minute. In addition, the patient had to meet at least one criterion in group 2, as follows: cardiovascular dysfunction, defined as a systolic blood pressure less than 90 mm Hg or a decrease in systolic pressure by at least 40 mm Hg for more than one

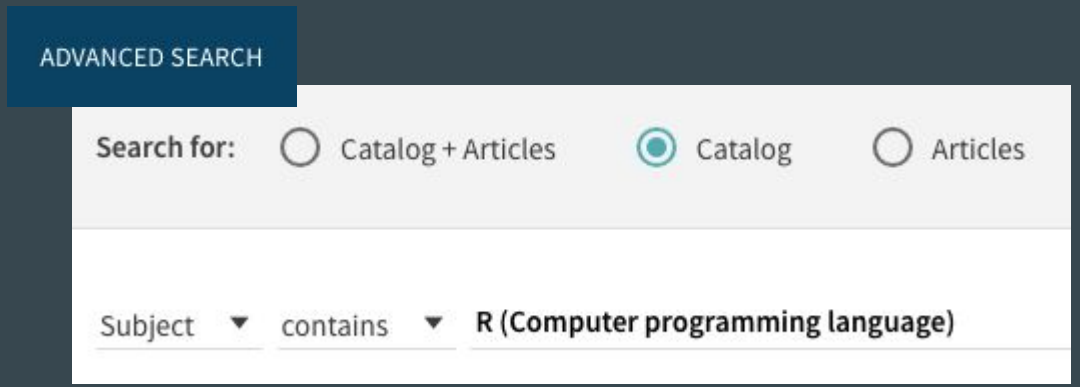
Some Handy R Links

Tutorials

- RStudio R paths: education.rstudio.com/learn
- Data Carpentry & Software Carpentry:
  - datacarpentry.org
  - software-carpentry.org
- Linkedin Learning @ GW: go.gwu.edu/linkedinlearning
- r-tutor.com/r-introduction & r-tutor.com/elementary-statistics
- UCLA Data Analysis Examples: stats.idre.ucla.edu/other/dae/
- R Graph Gallery (w/code): r-graph-gallery.com

Books you can access for free

- Free books online - Hadley Wickham:
  - R for Data Science r4ds.had.co.nz
  - Advanced R adv-r.hadley.nz/
- Through your GW library privileges:



Reference Links

- R language (CRAN): r-project.org
- R search engine: rseek.org
- rstudio.com
  - Cheat Sheets! rstudio.com/resources/cheatsheets
- stackoverflow.com

Statistics+R help @ GW

R-Statistics Appointments:
go.gwu.edu/data-consulting

Also...

Appointments with me:
calendly.com/kerchner or email me at kerchner@gwu.edu

Coding consultations (Python, R, git, etc.):
calendly.com/gwul-coding

Thanks!

Dan Kerchner kerchner@gwu.edu