Prove It! Reviewing the Medical Literature

Stephen C. Ausband, M.D. FACEP

Carilion Clinic

Plan

- Step through a fictitious study
- Definitions
- Basic Statistical concepts
- Types of scientific studies
- Process of medical research
- Critical approach to evaluating and utilizing information
- Interpreting and integrating research findings into practice
- Review studies

Caveat

- I didn't do all the math

$$\Pr_{\theta,\phi}(u(X) < \theta < v(X)) = \gamma \text{ for all } (\theta, \phi).$$

- I'm not advocating drinking beer and running calls

Our contribution to the literature

- Primary response fire based EMS agency
- Every day from 5p-9p, right at shift change we have a cookout with unlimited beer supply
- We notice our response times seem to be shorter from 5p-9p

Questions?

Definitions

- Hypothesis
  - proposed explanation of an observed phenomena
  - Falsifiable (useless to test an untestable hypothesis)
  - Parsimonious (fewest assumptions)
  - Scope (must apply to as much of the phenomena as possible)
  - Fit (should fit with prior knowledge)

Definitions

- Theory
  - Well substantiated
  - Repeatedly tested
  - Explanation of observed phenomena
  - Have predictive value
  - Provides explanations

Definitions

- Null hypothesis = there is no relation between intervention and outcome
- Type I error = null hypothesis incorrectly rejected
  - False positive result
- Type II error = null hypothesis incorrectly accepted
  - False negative result

Our contribution to the literature

- Hypothesis
  - Drinking beer shortens response times
- Null hypothesis
  - Drinking beer does not shorten response times
- Type I error = Drinking beer does not shorten response times but we incorrectly say it does (false positive)
- Type II error = Drinking beer does shorten response times, but we incorrectly say it doesn't (false negative)

Our contribution to the literature

- We start with the position that there is no relation between drinking the beer and response times
- We only accept that beer improves response times if the data are strong enough that we reject the null hypothesis

Definitions

- Data
  - Pieces of information
  - Values that can be reported, measured and analyzed

Definitions

- Statistics
  - Collection of data
  - Analysis of data
  - Interpretation of data
  - Organization of data

Definitions

- IRB
  - Internal Review board
  - Ensures safety of human subjects

Definitions

- Peer Review
  - Quality control process
  - Expert peers evaluate presented information
  - Blinded to the presenters

Definitions

- Sensitivity = ability of a test to pick up a condition
- Specificity = ability to correctly identify a condition exists
- Negative Predictive Value = chance that if a test says a condition isn't there, it really isn't there
- Positive predictive value = chance that if a test says a condition is there, it's really there

Definitions

- Number needed to treat = How many people need to be treated with something before one will have a positive effect
- Number needed to harm = How many people need to be treated with something before one will have a negative effect

Definitions

- Hawthorne effect
  - The act of observation changes subject behavior

Definitions

- Hawthorne effect example
- Hypothesis: Response times are shorter if EMS providers ingest > 2 beers before going on duty
- Method: have observers ride along and measure response times
- Results: response times are shorter with beer
- Problem: The same result occurs when any change is studied
- Reason: The providers know they are being observed. It's the act of observing that is the difference.

Questions?

Statistical concepts

- Mean = average
  - Is this useful?
  - The mean response time of our EMS agency is 9.8 min
  - 0, 0, 0, 0, 15, 12, 13, 16, 19, 23 min response times
  - I call 911, should I expect an ambulance at my door in < 10 minutes?

Statistical Concepts

- Standard deviation
  - How close is the data to the mean
  - Low standard deviation means the data is close to the mean
  - Higher standard deviations mean the data is farther from the mean

Standard Deviation

- The mean response time of our EMS agency is 9.8 min
- 0, 0, 0, 0, 15, 12, 13, 16, 19, 23 min response times
- SD = 8.9
- If I call 911, I can expect an ambulance in 9.8 +/- 8.9 min

Statistical Concepts

- P value
  - The chance that the differences in result are due to chance
  - Significance level set before analysis
    - If we set significance at 5% likelihood difference is due to chance we reject the null hypothesis if $p < 0.05$
  - If the P is < significance level, the results are unlikely due to chance
  - Why don't we set significance at 0% likelihood due to chance or 0.0000001% or something?

Statistical concepts

- Power
  - The test's probability of correctly rejecting the null hypothesis.
  - The chance you will find a significant effect
  - Calculated before data collection begins
  - No use starting a study if it can’t be powered adequately

Statistical concepts

- Hypothesis = toast lands preferentially on one side
- Null hypothesis = toast lands equally on one side or the other (significance 0.05, that is, we will reject the null hypothesis and accept our hypothesis if any difference we find is 5% or less likely due to random chance)
- Data = Butter up 14 times, butter down 6 times
- P = 0.12 (there is a 12% likely hood that the difference in side the toast landed on was due to chance)
- No significant difference in the results

Statistical concepts

- Confidence interval
  - The chance that repeated measurements will give results that include the true result
  - Everything in the confidence interval is a plausible result
  - Smaller intervals are better
  - If the confidence interval contains zero, the difference is not significant

Questions?

Our contribution to the literature

- Review what is known about response times
- Review what is known about beer and EMS, particularly regarding response times
- Refine hypothesis and define study parameters
  - Only beer? Any alcohol? What type of beer? Time between ingestion and going on duty, etc.
  - The more narrowly the hypothesis is defined, the better
  - The more narrowly the hypothesis is defined, the harder to study
- Logical, stepwise approach is best

Our contribution to the literature

- Study beer first
- If negative result, we can quit
- If positive result, we plan follow on study

Types of scientific studies

- Prospective = do something and measure
- Retrospective = something already happened, then measure

Types of scientific studies

- Randomized double blind placebo controlled
  - Prospective
  - Providers drink either 2 beers or 2 sham beers prior to shift
  - Who gets real beer and who doesn't and when is random
  - No one knows who's getting what or when
- Randomized placebo controlled
  - Prospective
  - Providers drink either 2 beers or 2 sham beers prior to shift
  - Who gets real beer and who doesn't and when is random
  - The providers (or researchers) don't know who's getting what or when

Types of scientific studies

- Case control
  - Prospective
  - One group of providers drinks 2 beers prior to shift
  - One group drinks sham beer or nothing
  - Both groups must be similar
- Before and after
  - Retrospective
  - In 2013 providers didn't drink beer before going on shift
  - In 2014 providers did drink beer before going on shift

Types of scientific studies

- Case review
  - Retrospective
  - Review cases where providers drank beer before going on shift vs cases where they didn't
- Case series
  - Retrospective
  - We had these calls after our weekly party and the call tiems were shorter
- Case report
  - One time we had this call after a party and our call time was shorter

Types of scientific studies

- Meta-analysis
  - Take several smaller studies and analyze the results together
  - Studies must be compatable

Quality of evidence

- Level A = high quality from more than 1 RCT, meta-analyses of high quality RCT
- Level B = moderate quality evidence from 1 or more RCT, meta-analyses of moderate quality RCT, well executed non-randomized trials,
- Level C = randomized or non-randomized observational studies with limitations, meta-analyses of same, expert consensus

Strength of recommendation

- Class I Benefit >>> risk
- Class IIa Benefit >> risk
- Class IIb Benefit > risk
- Class III No benefit
- Class III Harm

Real Literature Example

- "Routine aspiration thrombectomy before primary PCI is not useful
- Class III no benefit
- Level of evidence A
  - Levine GN, et al.
    2015 ACC/AHA/SCAI Focused Update on Primary PCI
  - ***Circulation*. 2015;000:000–000**

Questions?

Our contribution to the Literature

- How do we want to set this up?

Our Contribution to the literature

- All crews drink either a beer or placebo from 5p-9p (shift change at 5 pm)
- Drink is randomized and changed each day
- No one knows who is drinking what
- Outcome measure = En route time & on scene time for 1 year

Our contribution to the Literature

- We get IRB approval
- We follow all the right steps and procedures
- We do our study
- We analyze or data
- We submit to a peer reviewed journal
- We get feed back and re-submit

Our contribution to the literature

- Journal of Response and Repose, 2015 November, 15:10-12.
- **Beer Enhances Emergency Responses Trial.**
- **Simpson H, Flanders N, Bundy T**

Our Contribution to the literature

- **OBJECTIVE:**
- The objective of this study was to characterize changes in response time intervals in a busy rural EMS system when providers ingest 2 beers at the start of their work shift.

Our contribution to the literature

- **METHODS:**
- Anecdotal agency reports indicated that ingestion by providers of two or more beers in association with the weekly agency cook out was positively associated with improvement in time en route and time on scene. We endeavoured to evaluate for a causal relationship between beer ingestion at the begiining of a work shift by EMS providers and a decrease in en route and on scene times. All agency providers drank either 2 standard 12 oz beers (FizziPizz Light) at the beginning of their shift or a placebo which contained no alcohol. Assignement of either beer (B) or no beer (NB) each day was determined randomly by sealed number cards. All participants and investigators were blinded to assigned group. Time en route and time on scene intervals were analyzed and compared across a one year period.

Our contribution to the literature

- **RESULTS:**
- We identified a total of 5,403 transport events events during the one year study period. Mean time en route in the B group was 15.3 minutes (95% CI 14.7- 15.9 minutes). Mean time on scene in the B group was 12.9 minutes (95% CI 9.5 – 13.4 minutes). Mean time en route in the NB group was 16.4 minutes (95% CI 15.1- 17.9 minutes). Mean time on scene in the NB group was 13.3 minutes (95% CI 9.8 – 13.8 minutes). Differences in en route time was statistically significant ($P < 0.05$) as was the on scene time ($P < 0.05$)

Our Contribution to the literature

- **CONCLUSIONS:**
- Beginning of shift ingestion of 2 beers by EMS providers is associated with a positive improvement in en route and on scene times and should be strongly considered to be incorporated into regular practice.

Questions?

Questions?

- Lets look at a real study
- Keep in mind that it's usually an accumulation of knowledge / multiple studies that change general practice

Real literature

- Resuscitation. 2015 May;90:104-10.
- **Out-of-hospital pediatric airway management in the United States.**
- **Hansen M, Lambert W, Guise JM, Warden CR, Mann NC, Wang H**

Real Literature

- **OBJECTIVE:**
- The objective of this study was to characterize pediatric out-of-hospital airway management interventions, success rates, and complications in the United States using the 2012 National Emergency Medical Services Information System (NEMSIS) dataset.

Real Literature

- **METHODS:**
- In 2012, NEMSIS collected data from Emergency Medical Services (EMS) encounters in 40 states. We included all patients less than 18 years of age and identified all patients who had airway interventions including endotracheal intubation (ETI), bag-valve-mask ventilation (BVM), continuous positive airway pressure/bilevel positive airway pressure (CPAP/BiPAP) and alternate airways (Combitube, King LT, Laryngeal Mask Airway (LMA), esophageal obturator airway, and cricothyroidotomy). Success and complication rates were analyzed and compared across pediatric age groups, by race, ethnicity, clinical condition, and geographic region.

Real Literature

- **RESULTS:**
- We identified a total of 949,301 pediatric patient care events in the NEMSIS 2012 dataset. 4.5% had airway management procedures (42,936 events). Invasive airway management or ventilation (ETI, cricothyroidotomy, alternate airway, CPAP/BiPAP, BVM and other ventilation) took place in 1.5% of patient care events (14,107). Of those who had invasive airway management, 29.9% were less than 1 year of age, 58.1% were male, 42.3% were white, and 83.6% were in urban areas. ETI occurred in 3124 of patient care events (329 per 100,000; 95% CI 318-341). Overall success of ETI was 81.1% (95% CI 79.7-82.6). Lower success was noted in patients with cardiac arrest (75.5%, 95% CI 72.6-78.3) and those aged 1-12 months (72.1%, 95% CI 68.3-75.6).

Real Literature

- **CONCLUSIONS:**
- Out-of-hospital pediatric advanced airway procedures were infrequently performed. Success rates are lowest in patients aged 1-12 months.

Real Literature

- Discuss

Real Literature

- Prehosp Emerg Care. 2002 Oct-Dec;6(4):421-4.
- **Changes in physical examination caused by use of spinal immobilization.**
- **March JA, Ausband SC, Brown LH**

- **Abstract**
- The standard of care for patients following blunt trauma includes midline palpation of vertebrae to rule out fractures. Previous studies have demonstrated that spinal immobilization does cause discomfort.
- **OBJECTIVE:**
- To determine whether spinal immobilization causes changes in physical exam findings over time.

Real Literature

- **METHODS:**
- This was a single-blinded, prospective study at a tertiary care university teaching hospital. Twenty healthy volunteers without previous back pain or injuries, 13 male and seven female, were fully immobilized for one hour, with a cervical collar and strapped to a long wooden backboard. Midline palpation of vertebrae to illicit pain was performed at 10-minute intervals. In addition, the participants were asked to rate neck and back pain on a scale from 1 to 10 (1 for no pain, and 10 for unbearable pain), to see whether subjective pain from immobilization correlated with tenderness to palpation.

Real Literature

- **RESULTS:**
- Three patients had point tenderness of cervical vertebrae within 40 minutes. Five patients developed point tenderness of vertebrae by 60 minutes. Eighteen of 20 participants complained of increasing discomfort over time. The median initial pain scale was 1 (range 1-1), in contrast to 4 (range 1-9) at 60 minutes, $p < 0.05$.

Real Literature

- **CONCLUSION:**
- This study shows that over time, standard immobilization causes a false-positive exam for midline vertebral tenderness. In order to reduce this high false-positive rate for midline vertebral tenderness, the authors recommend that, initially on arrival to the emergency department, immediate evaluation occur of all immobilized patients. Furthermore, backboards should be modified to reduce patient discomfort to prevent the iatrogenically induced midline vertebral tenderness, thereby reducing subsequent false-positive examinations.

Discussion

Real Literature

- Is it peer reviewed? (published in a peer reviewed journal?)
- Does the hypothesis make sense?
- Does the set up and evaluation address the hypothesis question?

Real Literature

- Does the difference matter, really?
- Does this apply to you?
- Does this make sense?
- Is there some error or procedural problem?
- More studies to confirm / deny ?
- Useful study question? -> read abstract
- Abstract holds up? -> read the whole paper
- Always critically question, does is the hypothesis supported by data? If yes, even long held and cherished ideas may have to be reconsidered or rejected

Questions?