Clinical Epidemiology
Review For Step 1

Adam Obley
April 28, 2015
Outline

• Why this stuff matters (aside from Step 1)
• Study designs
• Type I and Type II errors
• Odds ratios
• Sensitivity, specificity, and LRAs
• Relative and absolute risk, NNT
• USMLE Questions
Why this matters...

- “Half” of what you have learned is wrong and potentially harmful to patients
- Pubmed has over 35 million citations and doubles in size every few years
- Certainty is illusion—nearly everything in medicine is just a matter of probabilities
- In the end, EBM is about honesty—with ourselves, our patients, and our students
A Decade of Reversal: An Analysis of 146 Contradicted Medical Practices

Vinay Prasad, MD; Andrae Vandross, MD; Caitlin Toomey, MD; Michael Cheung, MD; Jason Rho, MD; Steven Quinn, MD; Satish Jacob Chacko, MD; Durga Borkar, MD; Victor Gall, MD; Senthil Selvaraj, MD; Nancy Ho, MD; and Adam Cifu, MD

Abstract

Objective: To identify medical practices that offer no net benefits.

Methods: We reviewed all original articles published in 10 years (2001-2010) in one high-impact journal. Articles were classified on the basis of whether they addressed a medical practice, whether they tested a new or existing therapy, and whether results were positive or negative. Articles were then classified as 1 of 4 types: replacement, when a new practice surpasses standard of care; back to the drawing board, when a new practice is no better than current practice; reaffirmation, when an existing practice is found to be better than a lesser standard; and reversal, when an existing practice is found to be no better than a lesser therapy. This study was conducted from August 1, 2011, through October 31, 2012.

Results: We reviewed 2044 original articles, 1344 of which concerned a medical practice. Of these, 981 articles (73.0%) examined a new medical practice, whereas 363 (27.0%) tested an established practice. A total of 947 studies (70.5%) had positive findings, whereas 397 (29.5%) reached a negative conclusion. A total of 756 articles addressing a medical practice constituted replacement, 165 were back to the drawing board, 146 were medical reversals, 138 were reaffirmations, and 139 were inconclusive. Of the 363 articles testing standard of care, 146 (40.2%) reversed that practice, whereas 138 (38.0%) reaffirmed it.

Conclusion: The reversal of established medical practice is common and occurs across all classes of medical practice. This investigation sheds light on low-value practices and patterns of medical research.

Published by Elsevier Inc on behalf of Mayo Foundation for Medical Education and Research. Mayo Clin Proc. 2013;88(8):790-798
Study Designs

• Case series (including phase 1 and 2 trials)
• Cross-sectional studies
• Case-control studies
• Cohort studies
• Randomized controlled trials
• Systematic review and metaanalysis
Evidence Hierarchy

- Systematic reviews
- Critically-appraised topics [evidence syntheses and guidelines]
- Critically-appraised individual articles [article synopses]
- Randomized Controlled Trials (RCTs)
- Cohort studies
- Case-controlled studies case series / reports
- Background information / expert opinion
Context A relative paucity of data exist on the possible health effects of using cellular telephones.

Objective To test the hypothesis that using handheld cellular telephones is related to the risk of primary brain cancer.

Design and Setting Study conducted in 5 US academic medical centers between 1994 and 1998 using a structured questionnaire.

Patients A total of 469 men and women aged 18 to 80 years with primary brain cancer and 422 matched controls without brain cancer.

Main Outcome Measure Risk of brain cancer compared by use of handheld cellular telephones, in hours per month and years of use.

Results The median monthly hours of use were 2.5 for cases and 2.2 for controls. Compared with patients who never used handheld cellular telephones, the multivariate odds ratio (OR) associated with regular past or current use was 0.85 (95% confidence interval [CI], 0.6-1.2). The OR for infrequent users (<0.72 h/mo) was 1.0 (95% CI, 0.5-2.0) and for frequent users (>10.1 h/mo) was 0.7 (95% CI, 0.3-1.4). The mean duration of use was 2.8 years for cases and 2.7 years for controls; no association with brain cancer was observed according to duration of use ($P = .54$). In cases, cerebral tumors occurred more frequently on the same side of the head where cellular telephones had been used (26 vs 15 cases; $P = .06$), but in the cases with temporal lobe cancer a greater proportion of tumors occurred in the contralateral than ipsilateral side (9 vs 5 cases; $P = .33$). The OR was less than 1.0 for all histologic categories of brain cancer except for uncommon neuroepitheliomatous cancers (OR, 2.1; 95% CI, 0.9-4.7).

Conclusions Our data suggest that use of handheld cellular telephones is not associated with risk of brain cancer, but further studies are needed to account for longer induction periods, especially for slow-growing tumors with neuronal features.
Background—Despite the common coexistence of diabetes and heart failure (HF), the optimal medical treatment of diabetes in HF patients has not been well studied. We sought to compare the association between metformin use and clinical outcomes in a cohort of ambulatory patients with diabetes and established HF.

Methods and Results—Using propensity-score matched samples, we examined the association between metformin use and the risk of death or risk of hospitalization in a national cohort of 6185 patients with HF and diabetes treated in ambulatory clinics at Veteran Affairs medical centers. In this cohort, 1561 (25.2%) patients were treated with metformin. At two years of follow-up, death occurred in 246 (15.8%) patients receiving metformin and in 1177 (25.5%) patients not receiving metformin (p<0.001). In the propensity-score matched analysis (n=2874), death occurred in 232 (16.1%) patients receiving metformin compared to 285 (19.8%) patients not receiving metformin [hazard ratio (HR) 0.76, 95% CI 0.63 to 0.92, p< 0.01]. In propensity-score matched analyses, HF hospitalization or total hospitalization rates were not significantly different between individuals treated with metformin compared to those not treated with metformin (HR 0.93, 95% CI 0.74 to 1.18, and HR 0.94, 95% CI 0.83–1.07, respectively).

Conclusions—Metformin therapy was associated with lower rates of mortality in ambulatory patients with diabetes and HF. Future, prospective studies are necessary to define the optimal therapy for diabetic patients with HF.
Abstract

Severe sepsis in lung transplant recipients is a challenging problem and carries a high mortality. Recombinant human activated protein C (drotrecogin alfa [activated]) has been approved for use in patients with severe sepsis. Its use has been shown to be safe and impart a survival advantage. However, the safety of drotrecogin alfa activated has not been evaluated in lung transplant recipients. We report for the first time on the use of drotrecogin alfa activated in six lung transplant recipients. Clinical trials are warranted to further evaluate the use of drotrecogin alfa activated in transplant recipients.
Methods: A cross-sectional study was conducted to investigate the association between BPA exposure and body mass index (BMI) in school children. Three primary and three middle schools were randomly selected from 26 primary and 30 middle candidate schools in Changning District of Shanghai City in China. According to the BMI-based criteria by age and sex for screening of overweight or obese children, we randomly chose 20 obese, 10 overweight, and 30 normal weight children aged 8-15 years of age from each selected school. First morning urine was collected and total urine BPA concentrations were determined by ultra-performance liquid chromatography tandem mass spectrometry. Multiple linear regression analysis was conducted to examine the association of urine BPA concentrations and daily intake estimates with BMI.

Results: BPA was detected in 84.9% of urine samples with a geometric mean of 0.45 ng/mL. The daily intake estimates ranged from 0.03 μg/day to 1.96 μg/day with a geometric mean of 0.37 μg/day. The average urine BPA concentrations and daily intake estimates were similar for boys and girls, but significantly higher in older children than younger ones, and showed an increasing trend with BMI. Multiple linear regression analyses showed that urine BPA concentrations were significantly associated with increasing BMI values in all subjects after adjustment for age and sex and the results were similar before and after corrected by urine specific gravity. When stratified by age or sex, the associations remained significant in females and in those 8-11 years of age before corrected by specific gravity. Similar results were shown for the association between BMI and daily intake estimates.
METHODS
In this [REDACTED] trial, we [REDACTED] assigned 8442 patients with class II, III, or IV heart failure and an ejection fraction of 40% or less to receive either LCZ696 (at a dose of 200 mg twice daily) or enalapril (at a dose of 10 mg twice daily), in addition to recommended therapy. The primary outcome was a composite of death from cardiovascular causes or hospitalization for heart failure, but the trial was designed to detect a difference in the rates of death from cardiovascular causes.

RESULTS
The trial was stopped early, according to prespecified rules, after a median follow-up of 27 months, because the boundary for an overwhelming benefit with LCZ696 had been crossed. At the time of study closure, the primary outcome had occurred in 914 patients (21.8%) in the LCZ696 group and 1117 patients (26.5%) in the enalapril group (hazard ratio in the LCZ696 group, 0.80; 95% confidence interval [CI], 0.73 to 0.87; P<0.001). A total of 711 patients (17.0%) receiving LCZ696 and 835 patients (19.8%) receiving enalapril died (hazard ratio for death from any cause, 0.84; 95% CI, 0.76 to 0.93; P<0.001); of these patients, 558 (13.3%) and 693 (16.5%), respectively, died from cardiovascular causes (hazard ratio, 0.80; 95% CI, 0.71 to 0.89; P<0.001). As compared with enalapril, LCZ696 also reduced the risk of hospitalization for heart failure by 21% (P<0.001) and decreased the symptoms and physical limitations of heart failure (P=0.001). The LCZ696 group had higher proportions of patients with hypotension and nonserious angioedema but lower proportions with renal impairment, hyperkalemia, and cough than the enalapril group.
Intention-to-treat in RCTs

- Coronary Drug Project – 1980
- RCT of clofibrate (cholesterol drug) vs placebo in approx 5,000 men who had had a heart attack

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Mortality rate (5-year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clofibrate</td>
<td>20.0%</td>
</tr>
<tr>
<td>Placebo</td>
<td>20.9%</td>
</tr>
<tr>
<td>Clofibrate (&gt;80% adherence)</td>
<td>15.0%</td>
</tr>
<tr>
<td>Placebo (&gt;80% adherence)</td>
<td>15.1%</td>
</tr>
</tbody>
</table>
Type 1 and Type 2 Error

• Assume a hypothesis: Hot new drug X cures terrible disease Y
• Create the null: Hot new drug X has no effect on terrible disease Y
• Type 1 error occurs when we accept the hypothesis, but the null is actually true
• Type 2 error occurs when we accept the null, but the hypothesis is actually true
$\alpha$-statistic

- Is the predetermined threshold of statistical significance often given as the $p$-value.
- If a result is statistically significant at a $p$-value of $<0.05$ that means there is a $<5\%$ chance that our conclusion is due to chance alone.
- You could also phrase that as there is $<5\%$ chance of a type 1 error (accepting the hypothesis when the null is true).
- What would be the effect of setting the $\alpha$-statistic at $<0.01$ instead of $0.05$?
Odds ratios

Odds ratio calculation

\[
OR = \frac{a / b}{c / d} = \frac{ad}{bc}
\]

where

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>✔</td>
</tr>
<tr>
<td>c</td>
<td>✗</td>
</tr>
</tbody>
</table>

Example

\[
OR = \frac{354 / 143}{293 / 511}
\]

where

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔</td>
<td>354</td>
</tr>
<tr>
<td>✗</td>
<td>293</td>
</tr>
</tbody>
</table>

\[
OR = 4.32
\]
Diagnostic tests

• Assume a population of 100,000 people in which 1000 (1%) have lupus

• Propose universal screening with anti-dsDNA, at test with a sensitivity of 95% and specificity of 98%
Likelihood ratios

$$+LR = \frac{\text{sensitivity}}{1 - \text{specificity}} = 47$$

$$-LR = \frac{1 - \text{sensitivity}}{\text{specificity}} = 0.05$$
Applying a LR
Relative and Absolute Risk
A placebo-controlled clinical trial is conducted to assess whether a new antihypertensive drug is more effective than standard therapy. A total of 5000 patients with essential hypertension are enrolled and randomly assigned to one of two groups: 2500 patients receive the new drug and 2500 patients receive placebo. If the alpha is set at 0.01 instead of 0.05, which of the following is the most likely result?

(A) Significant findings can be reported with greater confidence
(B) The study will have more power
(C) There is a decreased likelihood of a Type II error
(D) There is an increased likelihood of statistically significant findings
(E) There is an increased likelihood of a Type I error
A new blood test to detect prostate cancer is evaluated in 300 male volunteers. A needle biopsy of the prostate gland is done on all men with serum prostate-specific antigen concentrations greater than 5 ng/mL (N<4). One hundred men undergo biopsy procedures; 90 are found to have prostate cancer, and five are found to have chronic prostatitis. Which of the following is necessary to calculate the sensitivity of this test?

(A) Incidence of chronic prostatitis in the general population
(B) Number of men with test results greater than 5 ng/mL and a normal biopsy specimen
(C) Prevalence of chronic prostatitis in the general population
(D) Prostate biopsies of men with test results equal to or below 5 ng/mL
An investigator is studying the effect of the number of hours watching television (Factor A) on the percent of hemoglobin A\textsubscript{1c} in people with type 2 diabetes mellitus. Two different variables, Factor A and hemoglobin A\textsubscript{1c}, are compared. The results of the study indicate a correlation coefficient of +0.9. Which of the following graphs shown best corresponds to these results?
Over 1 year, a study is conducted to assess the antileukemic activity of a new tyrosine kinase inhibitor in patients with chronic myeloid leukemia in blast crisis. All patients enrolled in the study are informed that they would be treated with the tyrosine kinase inhibitor. They are assigned to successive dose cohorts of 300 to 1000 mg/day of the drug. Six to eight patients are assigned to each dose. Treatment efficacy is determined based on the results of complete blood counts and bone marrow assessments conducted regularly throughout the study. This study is best described as which of the following?

(A) Case-control study
(B) Crossover study
(C) Open-labeled clinical trial
(D) Randomized clinical trial
(E) Single-blind, randomized, controlled trial
A population-based, case-control study is conducted to assess whether there is an association between statin use and incidence of colorectal cancer. A total of 900 participants are enrolled: 400 patients who were diagnosed with colorectal cancer between 1998 and 2004, and 500 healthy participants matched for age, sex, and ethnicity. A structured interview is used to determine statin use in the two groups. Results are shown:

<table>
<thead>
<tr>
<th>Statin Use</th>
<th>Colon Cancer</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
<td>100</td>
<td>200</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>300</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td>Total</td>
<td>400</td>
<td>500</td>
<td>900</td>
<td></td>
</tr>
</tbody>
</table>

Which of the following is the estimated odds ratio of colon cancer in statin-treated patients compared with patients with no statin treatment?

(A) -1.0
(B) -0.5
(C) 0
(D) 0.5
(E) 0.67
(F) 0.75
(G) 1.3
(H) 2
80. In a cohort study of elderly women, the relative risk ratio for hip fractures among those who exercise regularly is 1.2 (95% confidence interval of 1.1 to 1.8). Which of the following is the most appropriate conclusion about the effect of regular exercise on the risk for hip fracture?

(A) Statistically nonsignificant increase in risk
(B) Statistically nonsignificant overall decrease in risk
(C) Statistically significant overall decrease in risk
(D) Statistically significant overall increase in risk
A study is designed to evaluate the feasibility of acupuncture in children with chronic headaches. Sixty children with chronic headaches are recruited for the study. In addition to their usual therapy, all children are treated with acupuncture three times a week for 2 months. Which of the following best describes this study design?

(A) Case-control
(B) Case series
(C) Crossover
(D) Cross-sectional
(E) Historical cohort
(F) Randomized clinical trial
A study is conducted to assess the relationship between serum protein YY (PYY) concentrations and obesity. A total of 24 subjects with BMIs ranging from 17 to 40 kg/m² are enrolled in the study. A fasting serum PYY concentration is determined for each patient. The investigators report that the correlation coefficient between the two parameters is −0.84 (p<0.001). Based on these results, which of the following is the most accurate conclusion?

(A) A decreased serum PYY concentration is a cause of obesity

(B) The likelihood of a correlation between high BMIs and serum PYY concentrations is 0.16

(C) Obesity is not related to serum PYY concentrations

(D) The slope of the line showing the relationship between serum PYY concentrations and BMIs is −0.84

(E) Subjects with lower BMIs have higher serum PYY concentrations