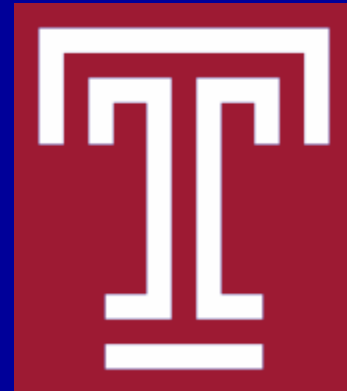


# Critiquing a Research Article

Temple DOM Research Curriculum

March 2015

Adam C. Ehrlich, MD, MPH  
Assistant Professor of Medicine  
Section of Gastroenterology



# Background

- The amount of medical knowledge is rapidly expanding.
- Even with the improvements in access that come with electronic indexing and the Internet, it is impossible to keep up with the medical literature.

# Background

- Some examples:
  - Pubmed search for “peptic ulcer disease” over only the last 5 years: 4,112 articles
  - Pubmed search for “constipation” over only the last 5 years: 5,629 articles
  - Pubmed search for “hepatitis C treatment” over only the last 5 years: 12,873 articles

# GI Journals (February 2015)

- Gastroenterology
  - 10 original articles, 5 editorials, 170 pages
- American Journal of Gastroenterology
  - 9 original articles, 12 editorials, 153 pages
- Clinical Gastroenterology and Hepatology
  - 20 original articles, 10 editorials, 191 pages
- Inflammatory Bowel Disease
  - 26 original articles, 242 pages
- Total for me:
  - 55 original articles, 27 editorials, 756 pages

# How do you get an article?

- You subscribe to the journal
- Someone gives you an article to read
- You search for an article to answer a clinical question
- You are required to review something for a journal club

# Background

Table 1—Ten reasons to read clinical journals

1. To impress others
2. To keep abreast of professional news
3. To understand pathobiology
4. To find out how a seasoned clinician handles a particular problem
5. To find out whether to use a new or existing diagnostic test on your patients\*
5. To learn the clinical features and course of a disorder\*
7. To determine etiology or causation\*\*
8. To distinguish useful from useless or even harmful therapy\*
9. To sort out claims concerning the need for and the use, quality and cost-effectiveness of clinical and other health care†
10. To be titillated by the letters to the editor

\*Reasons covered in detail in this series of Clinical Epidemiology Rounds.

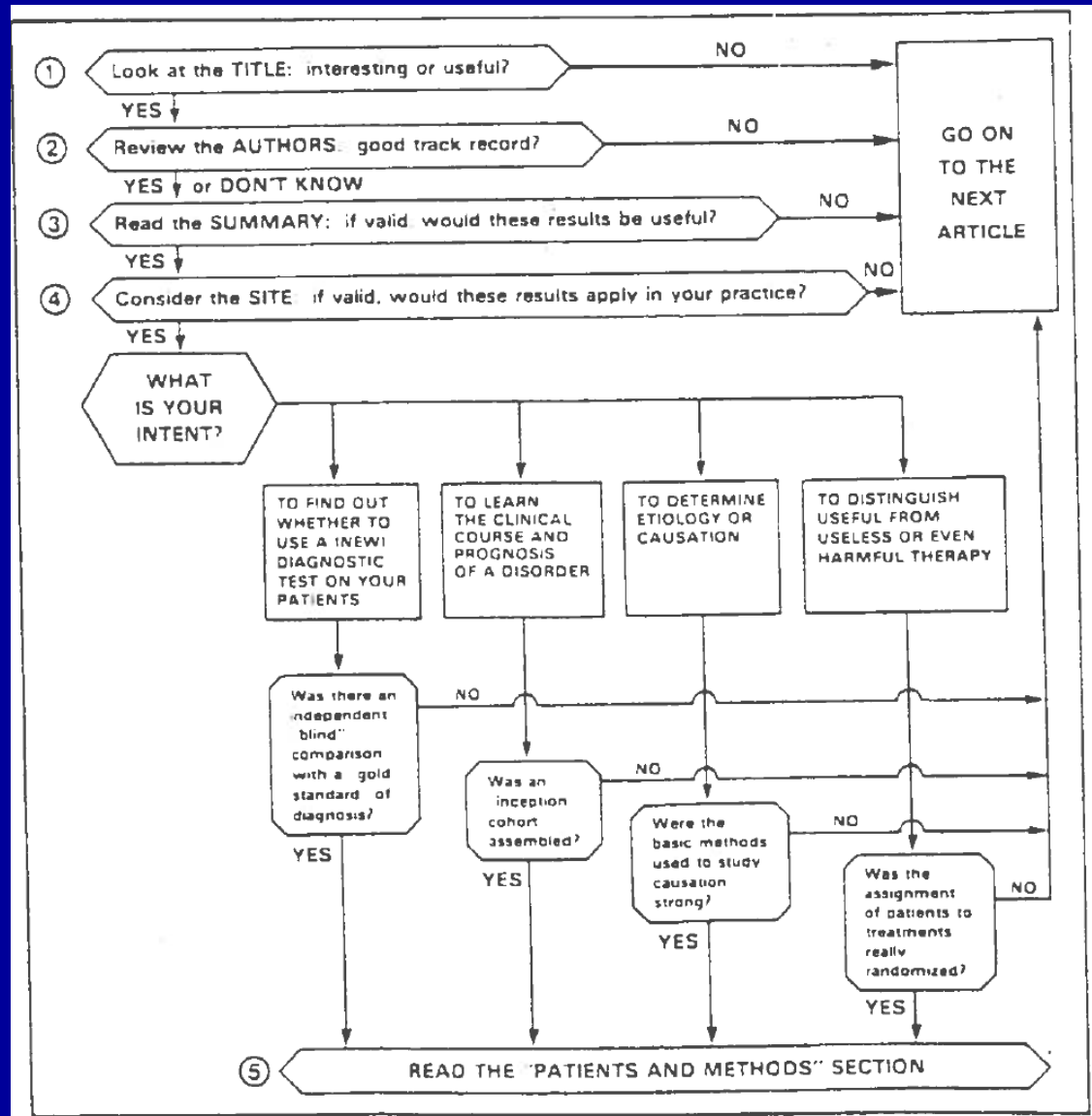
†A later series of Clinical Epidemiology Rounds will attempt to demystify the articles read for this reason.

# Disclaimer

- Remainder of the talk is my opinion
- There are a variety of methods and algorithms for doing a critical analysis of a journal article.
- Underlying theme is to be systematic.

# Beginning a critical analysis

- Initial review will not tell you if a paper is good, but it might tell you if it's bad
- My one addition - Journal





# Beginning a critical analysis

## The Risk for Cancer or Dysplasia in Ulcerative Colitis Patients With Primary Sclerosing Cholangitis <sup>1</sup>

Kirti Shetty, M.B.B.S., Lisa Rybicki, M.S., Aaron Brzezinski, M.D.,  
William D. Carey, M.D., F.A.C.G., and Bret A. Lashner, M.D., F.A.C.G. <sup>2</sup>

*Center for Inflammatory Bowel Disease, Departments of Gastroenterology and Biostatistics, Cleveland Clinic Foundation, Cleveland, Ohio* <sup>4</sup>

**OBJECTIVES:** Recent studies have implicated primary sclerosing cholangitis (PSC) as a risk factor for colorectal cancer (CRC) in ulcerative colitis (UC). Our study was designed to define both the risk and the risk factors for CRC or dysplasia in a large UC cohort with PSC.

**METHODS:** Patients with UC and PSC were compared with a random sample of UC controls without PSC. Patients were analyzed from the inception of disease until an outcome or censor.

**RESULTS:** Thirty-three (25%) of 132 UC patients with PSC developed CRC or dysplasia compared with 11 (5.6%) of 196 controls (adjusted relative risk 3.15, 95% confidence interval 1.37–7.27). Possible risk factors were chronic disease activity and lack of folate supplementation. Of 17 CRCs in the PSC group, 76% occurred proximal to the splenic flexure and 35% presented at an advanced stage, compared with one of five (20%) CRCs in controls being proximal and none being advanced. Six (4.5%) PSC patients, and no controls, died of CRC ( $p < 0.01$ ). <sup>3</sup>

**CONCLUSIONS:** UC patients with PSC are at increased risk of developing CRC or dysplasia. Chronically active disease may be a risk factor, whereas folate could have a protective effect. CRCs associated with PSC are more likely to be

5 – Prognosis study -> Was inception cohort assembled?

# OK, so now you're going to read the article...

- What is your purpose?
- Three major areas to assess:
  - Validity
  - Results
  - Applicability
- What type of article is it?
  - Cohort
  - Diagnosis
  - Prognosis
  - Treatment
  - Meta-analysis

# PICO

- Patients
- Intervention
- Comparator
- Outcome

# PICO

Are Your Patients Taking Their Medicine? Validation of a New Adherence Scale in Patients with Inflammatory Bowel Disease and Comparison with Physician Perception of Adherence

*Arvind J. Trindade, MD, Adam Ehrlich, MD, Asher Kornbluth, MD, and Thomas A. Ullman, MD*

- In (P) IBD patients, does (I) a new adherence scale (C) compared to physician perception and pill counts (O) predict adherence to medications?

# Validity

- Assessment of methodology of the study
- Dependent on the type of study
- Are there flaws that would render any observed significant differences less applicable?

# Results

- What are the results?
- Do the results represent a significant change or improvement from current treatments?

# Applicability

- Does the study apply my patient population? Were the study patients similar to my patients?
- Does the study apply to the particular patient I'm treating?
- If not, is it too great of a leap to apply the findings?

# Types of Studies



# The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812

OCTOBER 4, 2007

VOL. 357 NO. 14

## Effectiveness of Influenza Vaccine in the Community-Dwelling Elderly

Kristin L. Nichol, M.D., M.P.H., M.B.A., James D. Nordin, M.D., M.P.H., David B. Nelson, Ph.D.,  
John P. Mullooly, Ph.D., and Eelko Hak, Ph.D.

### ABSTRACT

#### BACKGROUND

Reliable estimates of the effectiveness of influenza vaccine among persons 65 years of age and older are important for informed vaccination policies and programs. Short-term studies may provide misleading pictures of long-term benefits, and residual confounding may have biased past results. This study examined the effectiveness of influenza vaccine in seniors over the long term while addressing potential bias and residual confounding in the results.

#### METHODS

Data were pooled from 18 cohorts of community-dwelling elderly members of one U.S. health maintenance organization (HMO) for 1990–1991 through 1999–2000 and of two other HMOs for 1996–1997 through 1999–2000. Logistic regression was used to estimate the effectiveness of the vaccine for the prevention of hospitalization for pneumonia or influenza and death after adjustment for important covariates. Additional analyses explored for evidence of bias and the potential effect of residual confounding.

#### RESULTS

There were 713,872 person-seasons of observation. Most high-risk medical conditions that were measured were more prevalent among vaccinated than among unvaccinated persons. Vaccination was associated with a 27% reduction in the risk of hospitalization for pneumonia or influenza (adjusted odds ratio, 0.73; 95% confidence interval [CI], 0.68 to 0.77) and a 48% reduction in the risk of death (adjusted odds ratio, 0.52; 95% CI, 0.50 to 0.55). Estimates were generally stable across age and risk subgroups. In the sensitivity analyses, we modeled the effect of a hypothetical unmeasured confounder that would have caused overestimation of vaccine effectiveness in the main analysis; vaccination was still associated with statistically significant — though lower — reductions in the risks of both hospitalization and death.

From the Medicine Service and Center for Chronic Disease Outcomes Research, Minneapolis Veterans Affairs Medical Center and University of Minnesota, Minneapolis (K.L.N., D.B.N.); the HealthPartners Research Foundation, Minneapolis (J.D.N.); Kaiser Permanente Northwest, Portland, OR (J.P.M.); and the Julius Center for Health Services and Primary Care, University Medical Center, Utrecht, the Netherlands (E.H.). Address reprint requests to Dr. Nichol at Medicine Service (111), VA Medical Center, 1 Veterans Dr., Minneapolis, MN 55417, or at nicho014@umn.edu.

N Engl J Med 2007;357:1373–81.

Copyright © 2007 Massachusetts Medical Society.

# Cohort Study

#### CONCLUSIONS

# Cohort Study

- Was the method for determining the exposure and control groups objective and accurate?
  - Consider causes of bias. Look for objective determinations of exposure and quantification of the amount of exposure.
- Were there any serious covariate imbalances?
  - Consider: There will always be differences, but the differences should not be so fundamental that comparisons do not make sense. Think of the major variables impacting the outcome of interest, and look for patient characteristics regarding those variables.

# Cohort Study

- Did the study adjust for important variables?
  - Consider: The authors should statistically adjust for any differences in important prognostic variables. Look for descriptions in the methods section of the adjustment process
- Is it unlikely that there were unmeasured differences between the groups that may have affected the outcome?
- Were all important outcomes considered?

# Comparison of Magnetic Resonance and Balloon Enteroscopic Examination of the Small Intestine in Patients With Crohn's Disease

Kento Takenaka,<sup>1</sup> Kazuo Ohtsuka,<sup>1</sup> Yoshio Kitazume,<sup>2</sup> Masakazu Nagahori,<sup>1</sup> Toshimitsu Fujii,<sup>1</sup> Eiko Saito,<sup>1</sup> Makoto Naganuma,<sup>1</sup> Akihiro Araki,<sup>1</sup> and Mamoru Watanabe<sup>1</sup>

<sup>1</sup>Department of Gastroenterology and Hepatology, <sup>2</sup>Department of Radiology, Tokyo Medical and Dental University, Tokyo, Japan

See Covering the Cover synopsis on page 257.

**BACKGROUND & AIMS:** Magnetic resonance (MR) enterography is a recommended imaging technique for detecting intestinal involvement in Crohn's disease (CD). However, the diagnostic accuracy of MR enterography has not been compared directly with that of enteroscopy of the jejunum and proximal ileum. We evaluated the usefulness of MR enterocolonography (MREC) by comparing its findings with those from balloon-assisted enteroscopy. **METHODS:** In a prospective study, MREC and enteroscopy were performed within 3 days of each other on 100 patients. Ulcerative lesions and all mucosal lesions were evaluated. Physicians and radiologists were blinded to results from other studies. Findings from MREC were compared directly with those from enteroscopy; the sensitivity and specificity with which MREC detected CD lesions were assessed. **RESULTS:** MREC detected ulcerative lesions and all mucosal lesions in the small intestine with 82.4% sensitivity (95% confidence interval [CI], 75.4%–87.7%) and 67.5% sensitivity (95% CI, 63.1%–70.0%); specificity values were 87.6% (95% CI, 83.7%–90.6%) and 94.8% (95% CI, 90.1%–97.5%). MREC detected major stenosis with 58.8% sensitivity (95% CI, 37.6%–77.2%) and 90.0% specificity (95% CI, 88.4%–91.5%) and all stenoses with 40.8% sensitivity (95% CI, 30.8%–49.4%) and 93.7% specificity (95% CI, 91.1%–95.9%). **CONCLUSIONS:** MREC is useful for detecting active lesions in the small intestine. However, MR imaging is less sensitive for detecting intestinal damage, such as stenoses. Enteroscopy is preferred for identifying intestinal damage. Suitable imaging approaches should be selected to assess CD lesions in deep small intestine.

**Keywords:** IBD; Diagnosis; Ulcer; Intestinal Damage.

natalizumab, and adalimumab.<sup>3–6</sup> Conventionally, imaging of CD lesions has relied mainly on ileocolonoscopy (ICS) and small-bowel follow-through (SBFT). ICS is useful for the detection of inflammation in the colon and the distal end of the ileum; however, it is impossible to access the deep small intestine using this procedure. Because small-bowel lesions are present in 4%–65% of CD patients,<sup>7–10</sup> conventional ICS has limitations in detecting CD lesions. SBFT is used to detect the presence of fistulae or mucosal damage in CD. However, the detection of small erosions or aphthae depends on the skill of the examiner. The range of diagnostic and therapeutic investigations for the small intestine in CD patients has widened considerably with recent technical advances such as wireless capsule endoscopy (WCE),<sup>11–13</sup> high-resolution computed tomography (CT),<sup>14</sup> and magnetic resonance (MR) enteroclysis or MR enterography (MRE).<sup>15–18</sup> Sensitivity values for the detection of extra-enteric complications in CD were significantly higher for CT and MRE than for SBFT.<sup>19</sup> In addition, balloon-assisted enteroscopy such as double-balloon endoscopy (DBE)<sup>20,21</sup> and single-balloon endoscopy (SBE)<sup>22</sup> are new techniques. An advantage of enteroscopy is that it enables concise assessment of the mucosa and acquisition of histopathologic specimens. In addition, endoscopic therapeutic procedures such as balloon dilatation of stenoses can be performed. Enteroscopy is expected to become an integral procedure for CD assessment.

The European Crohn's and Colitis Organisation guideline recommends MR or CT enterography or enteroclysis as imaging techniques with the highest diagnostic accuracy for the detection of intestinal involvement in CD, including extramural complications.<sup>23</sup> However, a recent study emphasized the high cumulative radiation dosages imparted

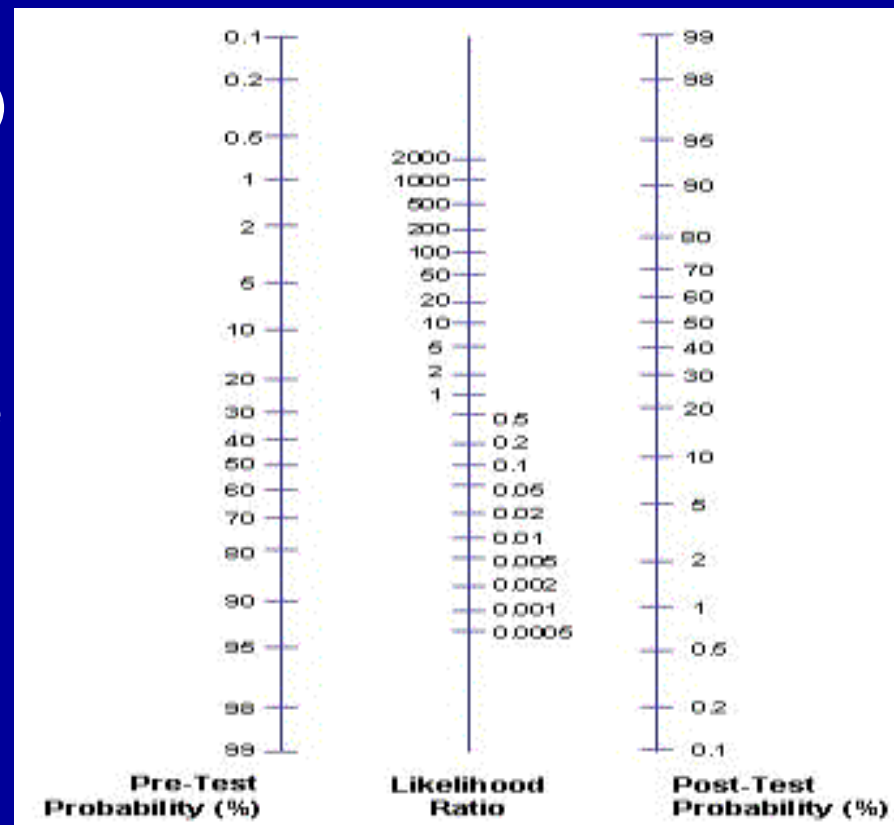
Diagnostic  
Test

# Diagnostic Test

- Did clinicians face diagnostic uncertainty?
  - Consider: Who are the patients and do they reflect the general population?
  - Did the patients need a diagnostic test or was the diagnosis clear?
- Was there a blind comparison between the test and an appropriate independent reference standard?
  - Blinding
  - Reference/Gold standard
- Was the gold standard test performed on everyone?

# Diagnostic Test

- What likelihood ratios were associated with the range of test results?
  - Describes the impact of the test result on the pre-test probability of the disease
  - $LR+ = \text{sensitivity}/(1-\text{specificity})$
  - $LR- = (1-\text{sensitivity})/\text{specificity}$
  - General guidelines:  $LR+ > 10$  essentially makes the diagnosis,  $LR- < 0.1$  essentially rules out the disease



# Diagnostic Test

- Is the test utilized in the study available?
- Are local providers capable of interpreting the test accurately?



# Prognosis

## The Risk for Cancer or Dysplasia in Ulcerative Colitis Patients With Primary Sclerosing Cholangitis

---

Kirti Shetty, M.B.B.S., Lisa Rybicki, M.S., Aaron Brzezinski, M.D.,  
William D. Carey, M.D., F.A.C.G., and Bret A. Lashner, M.D., F.A.C.G.

*Center for Inflammatory Bowel Disease, Departments of Gastroenterology and Biostatistics, Cleveland Clinic Foundation, Cleveland, Ohio*

**OBJECTIVES:** Recent studies have implicated primary sclerosing cholangitis (PSC) as a risk factor for colorectal cancer (CRC) in ulcerative colitis (UC). Our study was designed to define both the risk and the risk factors for CRC or dysplasia in a large UC cohort with PSC.

**METHODS:** Patients with UC and PSC were compared with a random sample of UC controls without PSC. Patients were analyzed from the inception of disease until an outcome or censor.

**RESULTS:** Thirty-three (25%) of 132 UC patients with PSC developed CRC or dysplasia compared with 11 (5.6%) of 196 controls (adjusted relative risk 3.15, 95% confidence interval 1.37–7.27). Possible risk factors were chronic disease activity and lack of folate supplementation. Of 17 CRCs in the PSC group, 76% occurred proximal to the splenic flexure and 35% presented at an advanced stage, compared with one of five (20%) CRCs in controls being proximal and none being advanced. Six (4.5%) PSC patients, and no controls, died of CRC ( $p < 0.01$ ).

**CONCLUSIONS:** UC patients with PSC are at increased risk of developing CRC or dysplasia. Chronically active disease may be a risk factor, whereas folate could have a protective effect. CRCs associated with PSC are more likely to be



# Prognosis

- Was the sample of patients appropriate to the question at hand and representative of patients with this problem?
  - Consider: Different forms of identifying the cohort may carry different biases. Referrals to a tertiary care center? Automatically enrolled?
- Were the patients sufficiently similar with respect to prognostic risk?
  - Are patients similar enough to analyze as a group?
- Was follow-up sufficiently long and complete?
- Was the primary outcome appropriate and clearly defined?

# The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812

DECEMBER 18, 2014

VOL. 371 NO. 25

## Cytisine versus Nicotine for Smoking Cessation

Natalie Walker, Ph.D., Colin Howe, Ph.D., Marewa Glover, Ph.D., Hayden McRobbie, M.B., Ch.B., Ph.D.,  
Joanne Barnes, Ph.D., Vili Nosa, Ph.D., Varsha Parag, M.Sc., Bruce Bassett, B.A.,  
and Christopher Bullen, M.B., Ch.B., Ph.D.

### ABSTRACT

#### BACKGROUND

Placebo-controlled trials indicate that cytisine, a partial agonist that binds the nicotinic acetylcholine receptor and is used for smoking cessation, almost doubles the chances of quitting at 6 months. We investigated whether cytisine was at least as effective as nicotine-replacement therapy in helping smokers to quit.

#### METHODS

We conducted a pragmatic, open-label, noninferiority trial in New Zealand in which 1310 adult daily smokers who were motivated to quit and called the national quitline were randomly assigned in a 1:1 ratio to receive cytisine for 25 days or nicotine-replacement therapy for 8 weeks. Cytisine was provided by mail, free of charge, and nicotine-replacement therapy was provided through vouchers for low-cost patches along with gum or lozenges. Low-intensity, telephone-delivered behavioral support was provided to both groups through the quitline. The primary outcome was self-reported continuous abstinence at 1 month.

#### RESULTS

At 1 month, continuous abstinence from smoking was reported for 40% of participants receiving cytisine (264 of 655) and 31% of participants receiving nicotine-replacement therapy (203 of 655), for a difference of 9.3 percentage points (95% confidence interval, 4.2 to 14.5). The effectiveness of cytisine for continuous abstinence was superior to that of nicotine-replacement therapy at 1 week, 2 months, and 6 months. In a prespecified subgroup analysis of the primary outcome, cytisine was superior to nicotine-replacement therapy among women and noninferior among men. Self-reported adverse events over 6 months occurred more frequently in the cytisine group (288 events among 204 participants) than in the group receiving nicotine-replacement therapy (174 events among 134 participants); adverse events were primarily nausea and vomiting and sleep disorders.

#### CONCLUSIONS

When combined with brief behavioral support, cytisine was found to be superior to nicotine-replacement therapy in helping smokers quit smoking, but it was as

From the National Institute for Health Innovation (N.W., C.H., V.P., C.B.), the Centre for Tobacco Control Research, Department of Social and Community Health (M.G.), the Department of Pacific Health (V.N.) and the School of Population Health and the School of Pharmacy (J.B.), University of Auckland, Auckland, and the Quit Group, Wellington (B.B.) — all in New Zealand; and Queen Mary University of London, Wolfson Institute of Preventive Medicine, and UK Centre for Tobacco and Alcohol Studies, Barts and the London School of Medicine and Dentistry — all in London (H.M.). Address reprint requests to Dr. Walker at the National Institute for Health Innovation, School of Population Health, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand; or at n.walker@auckland.ac.nz.

*N Engl J Med* 2014;371:2353-62.

DOI: 10.1056/NEJMoa1407764

Copyright © 2014 Massachusetts Medical Society.

# Treatment

# Treatment

- Was the assignment of patients to treatments randomized? Was allocation concealed?
- Were all patients who entered the trial accounted for at its conclusion?
  - How and why were patients lost to follow up?
- Were patients analyzed in the groups to which they were randomized?
  - Intention to treat, modified intention to treat, per protocol
- Were patients and clinicians kept blind to which treatment was being received?
  - Patients, doctors, outcome assessors
- Were the groups similar at the start of the trial?
  - Table 1

# Treatment

- Number needed to treat –  $1/\text{absolute risk reduction}$
- Absolute risk reduction = control event rate – experimental event rate

Therapy	Event prevented	Length of follow-up	NNT
H. pylori eradication in duodenal ulcer	Ulcer recurrence	1 year	1.1
Finasteride for benign prostatic hypertrophy	Need for surgery	2 years	39
Streptokinase and aspirin for acute myocardial infarction	One death	5 weeks	20
Enalapril for mild/moderate heart failure	One death	1 year	100
Lipid lowering in patients with coronary heart disease	One myocardial infarction (MI) or stroke related death	5 years	16
Treatment of mild high blood pressure	One MI, stroke or death.	1 year	700
Treatment of severe hypertension	One MI, stroke or death	1 year	15

# Meta-Analysis

## Thrombolysis Compared With Heparin for the Initial Treatment of Pulmonary Embolism A Meta-Analysis of the Randomized Controlled Trials

Susan Wan; Daniel J. Quinlan, MBBS; Giancarlo Agnelli, MD; John W. Eikelboom, MBBS

**Background**—Randomized trials and meta-analyses have reached conflicting conclusions about the role of thrombolytic therapy for the treatment of acute pulmonary embolism.

**Methods and Results**—We performed a meta-analysis of all randomized trials comparing thrombolytic therapy with heparin in patients with acute pulmonary embolism. Eleven trials, involving 748 patients, were included. Compared with heparin, thrombolytic therapy was associated with a nonsignificant reduction in recurrent pulmonary embolism or death (6.7% versus 9.6%; OR 0.67, 95% CI 0.40 to 1.12,  $P$  for heterogeneity=0.48), a nonsignificant increase in major bleeding (9.1% versus 6.1%; OR 1.42, 95% CI 0.81 to 2.46), and a significant increase in nonmajor bleeding (22.7% versus 10.0%; OR 2.63, 95% CI 1.53 to 4.54; number needed to harm=8). Thrombolytic therapy compared with heparin was associated with a significant reduction in recurrent pulmonary embolism or death in trials that also enrolled patients with major (hemodynamically unstable) pulmonary embolism (9.4% versus 19.0%; OR 0.45, 95% CI 0.22 to 0.92; number needed to treat=10) but not in trials that excluded these patients (5.3% versus 4.8%; OR 1.07, 95% CI 0.50 to 2.30), with significant heterogeneity between these 2 groups of trials ( $P=0.10$ ).

**Conclusions**—Currently available data provide no evidence for a benefit of thrombolytic therapy compared with heparin for the initial treatment of unselected patients with acute pulmonary embolism. A benefit is suggested in those at highest risk of recurrence or death. The number of patients enrolled in randomized trials to date is modest, and further evaluation of the efficacy and safety of thrombolytic therapy for the treatment of high-risk patients with acute pulmonary embolism appears warranted. (*Circulation*. 2004;110:744-749.)

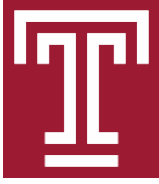
Key Words: embolism ■ meta-analysis ■ thrombolysis ■ heparin

# Meta-Analysis

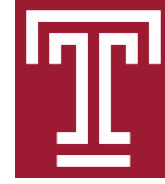
- Did the review ask a clear and focused clinical question?
- Did the review include the right type of article?
  - RCTs, other articles
  - Did the included studies address the question or is the data nested in another study?
- Were all relevant studies identified?
  - Databases searched, search terms, reference tracking, unpublished studies, non-English
- Was the validity of the included studies appraised?
  - Garbage in, garbage out
- Were assessments of study quality reproducible?
  - Agreement between assessors (Kappa statistic)
- Were results combined appropriately?
  - Test for heterogeneity –  $I^2 < 50\%$  is acceptable or using Random Effects Model

# Summary

- Critiquing a journal article is an important skill both for academic and private practice physicians
- After deciding to read an article, the key is to be systematic
- Depending on study type, look for methodological validity, results, and applicability
- I have “cheat sheets” for reviewing different types of articles if you are interested.



## Critical Appraisal Form - Cohort studies



This critical appraisal form should be used for cohort studies about prognosis or benefit or harm of a treatment or exposure. The following questions will help focus your attention on the important methodological issues related to cohort studies. They are divided into three sections: validity, results, applicability.

### VALIDITY:

#### 1. Was the method for determining the exposure and control groups objective and accurate? How was the control group established?

Consider:

- Different forms of identifying cohorts or exposure within them may carry different degrees of bias. Patient report of exposures may have associated recall bias.
- In general, look for objective determinations of exposure, and quantification of the amount of exposure.

Yes           Cannot tell           No

How was the cohort established? How were the exposure and control groups established?

---

---

---

#### 2. Were there any serious covariate imbalances? In other words, were the two groups adequately similar at the start of the trial?

Consider:

- There will always be differences, but the differences should not be so fundamental that comparisons do not make sense
- Think of the major variables impacting the outcome of interest, and look for patient characteristics regarding those variables.

Yes           Cannot tell           No

What were the important covariates (for the outcome of interest) and were they similar?

---

---



**3. Did the study adjust for important variables?**

Consider:

- The authors should statistically adjust for any differences in important prognostic variables.
- Look for descriptions in the methods section of the adjustment process

Yes                       Cannot tell                       No

What variables did the authors adjust for? Do these seem appropriate?

---

---

---

**4. Is it unlikely that there were unmeasured differences between the groups that may have affected the outcome?**

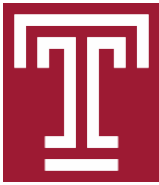
**5. Were all important outcomes considered?**

**Results:**

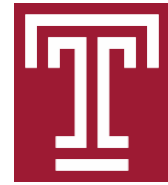
- **What was the adjusted hazard ratio for the primary outcome?**

**Applicability:**

- **Can you apply these results to your patients? Were the included patients similar to your own patient(s)?**



## Critical Appraisal Worksheet – Diagnostic Test



The following questions will help focus your attention on the important methodological issues related to articles on diagnosis. They are divided into three sections: validity, results, applicability.

### VALIDITY

#### 1. Did clinicians face diagnostic uncertainty?

Think about:

- Who are the patients? Do they reflect a general population or a biased one?
- Did all the patients actually need a diagnostic test (or was the diagnosis already clear)?
- The spectrum of disease, which includes issues of disease severity and alternate diagnoses in the study patients. Spectrum can have a large impact on measures of sensitivity and specificity. The appropriate spectrum of patients for a study should reflect patients who would receive the test in real life.

Yes                       Cannot tell                       No

What type of patient was enrolled in the study?

---



---



---

#### 2. Was there a blind comparison between the test and an appropriate independent reference standard?

Think about:

- Blinding - The people interpreting the reference standard should be unaware of the result of the test being studied, and people interpreting the test under study should be unaware of the reference standard results.
- What is the reference/gold standard? Is it reasonable? Can you think of a better one?
- Keep in mind that sometimes a study might use a complex reference standard, in which diagnosis may be established in different ways (e.g. biopsy OR long-term follow up)

Yes                       Cannot tell                       No

Describe the reference standard and whether it is an appropriate one:

---



---



---

### 3. Was the reference (“gold”) standard test performed on all patients regardless of the result of the test being evaluated?

Think about

- Whether patients with both positive and negative index tests actually received the reference standard test
- If not all patients receive the reference standard, *verification bias* is said to be present

Yes                       Cannot tell                       No

Describe which patients received the reference standard:

---

---

---

## RESULTS

### What likelihood ratios were associated with the range of possible test results?

Consider:

- *Likelihood ratios* (LR) describe the impact of a test result on the pre-test probability of disease and can be calculated for a positive test, a negative test, or for a particular test result or range of results.
- The LR is the ratio of likelihood of having disease with a given test result divided by the likelihood of not having disease with that same test result
- The LR for a positive test is: ***sensitivity/1-specificity*** and for a negative test is ***1-sensitivity/specificity***
- A LR of >10 for a positive test means that a positive result essentially “makes the diagnosis”; a negative LR of <.1 means a negative test can essentially rule out the disease.
- See the end of this document for a figure to apply the likelihood ratio to the pre-test probability to establish a post-test probability of having the disease.

What are the likelihood ratios?

---

---

---

---

## APPLICABILITY

### 1. Will the reproducibility of the test result and its interpretation be satisfactory in my clinical setting?

Consider

- Is the test utilized in the study available?
- Are local providers capable of interpreting the test accurately?

Yes                       Cannot tell                       No

Describe any issues related to using the test locally:

---

---

## 2. Are the results applicable to the patient in my practice?

Consider if the patients in the study are adequately similar to my own patients, and in particular to patients in whom I would order this test.

Yes                       Cannot tell                       No

Why (or why not) are the results applicable to your patients? What aspects of patient spectrum contribute to applicability?

---

---

## 3. Will the results change my management strategy?

Consider

- By how much do the LRs associated with this test change my pre-test probability of disease?
- Is the test accurate enough to impact my treatment plan? In what type of patients is it useful?

Yes                       Cannot tell                       No

How would this test change management strategy?

---

---

---

## 4. Will patients be better off as a result of the test?

Consider

- In what way will patients be better off?
- Will this test spare patients from having other testing done?
- How do we really determine if patients are better off?

Yes                       Cannot tell                       No

Describe how patients will be better off as a result of this test.

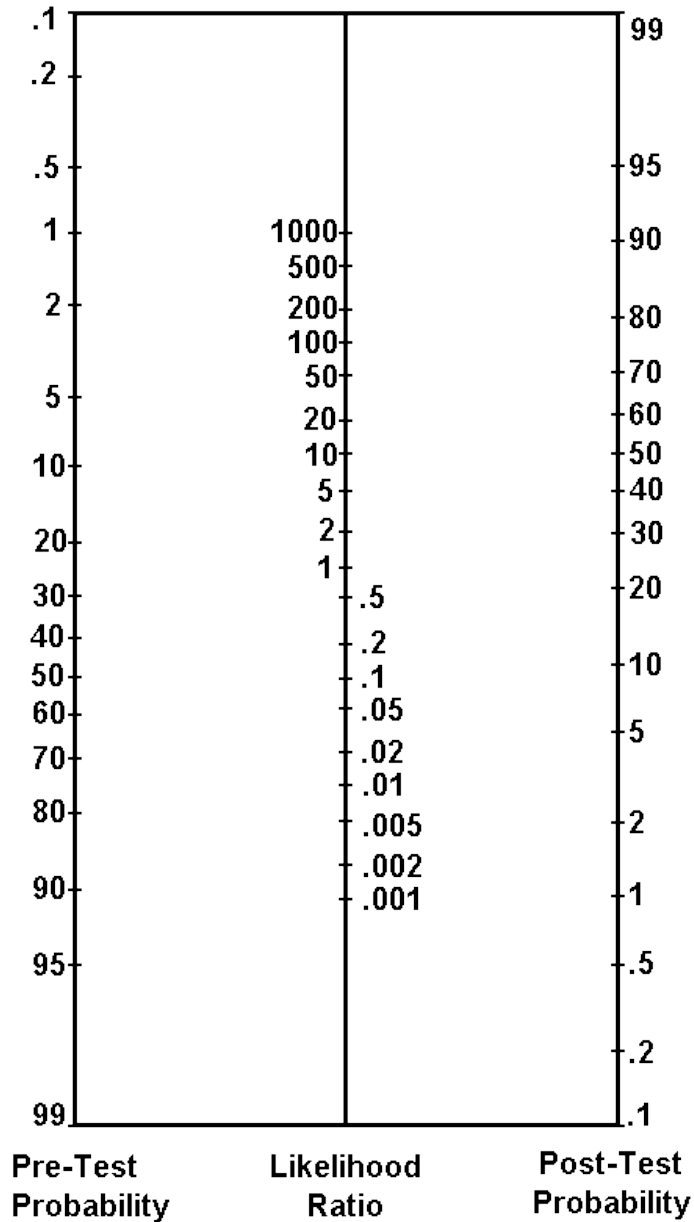
---

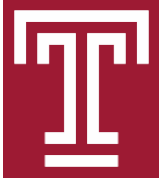
---

---

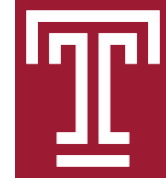
# Nomogram for Likelihood Ratio

1. Estimate the pre-test probability. This may be from previously established data or your estimate that is specific to your patient.
2. Calculate the likelihood ratio as described above in the results section.
3. Draw a straight line between the pre-test probability and the likelihood ratio. If you continue the line, you can connect it to the post-test probability.





## Critical Appraisal Worksheet for Meta-Analysis



The following questions will help focus your attention on the important methodological issues related to the systematic review. They are divided into three sections: validity, results and applicability. This document is an adaptation of the usual critical appraisal tool for systematic reviews.

### VALIDITY

#### 1. Did the review ask a clear and focused clinical question?

Consider if the question is focused in terms of

- The population studied (in terms of their risk of disease, co-morbidities, setting, etc)
- The intervention
- The outcomes considered

Yes

Cannot tell

No

What was the clinical question?

---

---

#### 2. Did the review include the right type of study?

Consider the criteria for study inclusion in terms of

- Study design- did they include only RCTs, or other articles as well?
- Whether the included studies addressed the clinical question at hand

Yes

Cannot tell

No

Describe the criteria for article inclusion.

---

---

#### 3. Were all relevant studies identified?

Consider the following issues:

- Which databases were used?
- What were the search terms?
- Was there follow-up from references? (known as “reference tracking”)
- Were unpublished studies identified?
- Was publication bias considered? (Look for mention of a funnel plot)
- Did the reviewers consider non-English language studies?

Adam C. Ehrlich, MD, MPH

Department of Medicine, Section of Gastroenterology

Temple University School of Medicine

Modified from Deborah Korenstein, MD at Icahn School of Medicine at Mount Sinai

Yes                       Cannot tell                       No

Describe issues related to study identification:

---

---

#### 4. Was the validity of the included studies appraised?

The authors should assess quality issues related to studies of diagnostic tests

- Look for a described scoring system
- Many studies will use the JADAD score, which is one generally accepted standard but there are other methods for scoring quality – See table at end
- Quality issues for the included studies:
  - Randomization
  - Allocation concealment
  - Blinding
  - Loss to follow up

Yes                       Cannot tell                       No

If yes, summarize the overall quality of the included studies.

---

---

#### 5. Were assessments of study quality reproducible?

Look for

- More than one assessor of study quality
- A report of the agreement of these assessors and how disagreements were resolved
- A kappa statistic, which describes the agreement of the assessors **beyond** chance agreement. Kappa ranges from 0 (no agreement) to 1 (perfect agreement). A kappa above .5 or .6 is considered acceptable and above .8 is considered excellent.

Yes                       Cannot tell                       No

Describe the study quality assessment process, and the kappa, if given:

---

---

## 6. If results were combined, was it done appropriately?

Consider the following issues:

- Results should only be combined if they are similar from study to study.
- Results of each study should be presented clearly, usually in a Forest plot showing the effect of the intervention in each study. Sometimes these are included in supplementary materials.
- Were the results similar from study to study? Look for a test of heterogeneity (or homogeneity). The favored test for heterogeneity is the  $I^2$  test. The result ranges from 0-100%, with 0% indicating no heterogeneity (i.e. all differences in study results are due to chance alone) and 100% indicating heterogeneity (i.e. no differences in study results are due to chance alone). In general,  $I^2$  below 50% represents acceptable lack of heterogeneity, meaning that the studies are similar and pooling results is appropriate.
- If there is no heterogeneity, authors will generally pool results using the Fixed Effects Model. If there is heterogeneity, look for them to use the Random Effects Model, which is more conservative and allows for pooling of somewhat disparate studies, with wider confidence intervals.

Yes

Cannot tell

No

Describe the degree of similarity among the studies and how pooling was done.

---

---

---

## RESULTS

### 1. How are the results presented and what are the main results?

Look for

- How results are presented- usually RR or OR are combined for the main result

Describe the main result and how it is presented.

---

---

---

### 2. How precise are the results?

Look for

- Confidence intervals around the main summary results
- Does the confidence interval cross 1?

Adam C. Ehrlich, MD, MPH

Department of Medicine, Section of Gastroenterology

Temple University School of Medicine

Modified from Deborah Korenstein, MD at Icahn School of Medicine at Mount Sinai



Is precision addressed, and if so, how?

---

---

---

## APPLICABILITY

### 1. Were all clinically important outcomes considered?

Yes                       Cannot tell                       No

What are the most important outcomes?

---

---

### 2. Can the results be applied to my patient care?

Consider:

- How does the patient population compare to my own?
- Were the interventions similar to what I can accomplish in my setting
- The overall quality of the included studies.

How does this review apply to patient care?

---

---

---

# Modified JADAD score

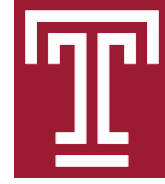
≥4 considered good quality study

Table 1	
Modified Jadad score	
Item Assessed	Score
Was the study described as randomized?	
Yes	+1
No	0
Was the method of randomization appropriate?	
Yes	+1
No	-1
Not described	0
Was the study described as blinded?	
Yes	+1
No	0
Was the method of blinding appropriate?	
Yes	+1
No	-1
Not described	0
Was there a description of withdrawals and dropouts?	
Yes	+1
No	0
Was there a clear description of the inclusion/exclusion criteria?	
Yes	+1
No	0
Was the method used to assess adverse effects described?	
Yes	+1
No	0
Was the method of statistical analysis described?	
Yes	+1
No	0

Adam C. Ehrlich, MD, MPH  
 Department of Medicine, Section of Gastroenterology  
 Temple University School of Medicine  
 Modified from Deborah Korenstein, MD at Icahn School of Medicine at Mount Sinai



## Critical Appraisal Form - Prognosis



This critical appraisal form should be used for studies of prognosis. Studies of prognosis are generally cohort studies, in which the cohort or part of a cohort is followed over time to determine prognosis. The following questions will help focus your attention on the important methodological issues related to cohort studies. They are divided into three sections: validity, results, applicability.

### VALIDITY:

#### 1. Was the sample of patients appropriate to the question at hand and representative of patients with this problem?

Consider:

- Different forms of identifying cohorts may carry different degrees of bias. Consider if patients pass through any filters before being enrolled in the study, such as referral to a specialist or tertiary center, or whether patients were enrolled automatically or referred into the study.
- Look for a concrete definition of health or disease status and severity.
- The original group of patients identified is called the “inception cohort”

Yes                       Cannot tell                       No

How were patients identified?

---



---



---

#### 2. Were the patients sufficiently similar with respect to prognostic risk?

Consider:

- Are the patients similar enough that it makes sense to analyze them as a group?
- Think of the major variables impacting the outcome of interest, and look for patient characteristics regarding those variables.

Yes                       Cannot tell                       No

What were the important covariates (i.e. determinants of the outcome of interest) and were they homogeneous within the group?

---



---

### 3. Was follow up sufficiently long and complete?

Consider

- Was the follow up time adequate for the outcome of interest?
- Were many patients lost to follow up? If so, how might this impact the findings?

Yes           Cannot tell           No

Describe any issues related to follow up:

---

---

### 4. Was the primary outcome appropriate and clearly defined?

Consider

- Was the primary outcome the most important outcome measure or was it a surrogate marker?
- How was the primary outcome defined? Look for objective and unbiased outcome criteria.

What are the most important outcomes for the issue at hand? What was the primary outcome of the study?

---

---

---

### Results:

#### • How likely are the outcomes over time?

The main result may be a hazard ratio or a relative risk, or it may be the likelihood of the outcome over time, represented by a survival curve. Look for it to be **adjusted**, after multivariate analysis, which accounts for confounders. Think of a hazard ratio essentially the way you would think about a relative risk.

Write the adjusted relative risk or hazard ratio for the primary outcome or describe the results of the survival curve, and explain what this means about prognosis.

---

---

---

### Applicability:

#### • Can you apply these results to your patients? Were the included patients and their management similar to your own patient(s)?

Think about

- The properties of the included patients and whether those properties make it different from other populations
- Was follow up sufficiently long?
- Whether confounding factors were adequately adjusted for

Yes

Cannot tell

No

What are the important issues regarding the applicability of this study?

---

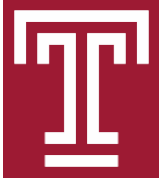
---

Adam C. Ehrlich, MD, MPH

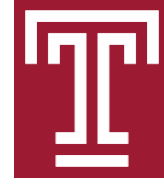
Department of Medicine, Section of Gastroenterology

Temple University School of Medicine

Modified from Deborah Korenstein, MD at Icahn School of Medicine at Mount Sinai



## Critical Appraisal Worksheet - Treatment



The following questions will help focus your attention on the important methodological issues related to articles on therapy. They are divided into three sections: validity, results, applicability.

### VALIDITY

#### 1. Was the assignment of patients to treatments randomized? Was allocation concealed?

Consider:

- How was the randomization done? Is that method likely to be effective?
- *Allocation concealment* means that investigators assessing patients for entry into the trial (PRIOR to randomization) would be unable to predict to which group the next patient will be randomized. Authors should describe the method for randomization (centrally done, opaque envelopes) in sufficient detail to ascertain allocation concealment.

Yes                       Cannot tell                       No

Describe the method of randomization and allocation concealment:

---

---

---

#### 2. Were all patients who entered the trial accounted for at its conclusion?

Consider:

- How many patients were lost to follow up?
- Why patients were lost to follow up?
- How the study considered patients who were lost> Many studies will carry forward the last available data for patients who were lost to follow up.

Yes                       Cannot tell                       No

How complete was the follow up and how did the investigators manage the data from patients who dropped out?

---

---

#### 3. Were patients analyzed in the groups to which they were randomized?

Adam C. Ehrlich, MD, MPH

Department of Medicine, Section of Gastroenterology

Temple University School of Medicine

Modified from Deborah Korenstein, MD at Icahn School of Medicine at Mount Sinai

This is known as *intention-to-treat* analysis, in which patients are counted in their original group regardless of the treatment they ultimately receive. It is important in minimizing bias. Sometimes when many patients “crossed over” to the other group investigators will also do a *per protocol*, or *as treated*, analysis, which looks at the treatment each patient actually received.

Yes                       Cannot tell                       No

How did the investigators manage cross overs?

---

---

---

#### 4. Were patients and clinicians kept blind to which treatment was being received?

Consider all the parties who should be blinded:

- Patient receiving the treatment, especially if outcomes are more subjective
- Doctors caring for the patients
- Outcome assessors. This is probably the most important. In studies in which patients cannot be blinded for logistical reasons, the investigators determining which patients met study endpoints should still be blinded
- Investigators writing the manuscript (until the last minute). Few studies do this but it probably represents the least biased method for interpreting the data.

Yes                       Cannot tell                       No

Who was blinded? If aspects of the study were not blinded, how much bias was introduced?

---

---

---

#### 5. Were the groups similar at the start of the trial?

Consider

- This information is often presented in “Table 1”
- Overall, were the groups similar with regard to most features?
- Are all the pertinent features of the patient population presented? The paper should describe patient similarity with regard to all variables that might impact the outcome of interest.
- If there are differences between the groups, are they clinically (as opposed to statistically) significant? In which direction do the differences bias the study (i.e. in favor of which group doing better)?

Yes                       Cannot tell                       No

In what ways did the groups differ? List any clinically important variables that were not described.

---

---

## RESULTS

Results are often presented primarily as a *Number Needed to Treat*, or **NNT**. The NNT is 1/Absolute Risk Reduction (expressed as a decimal). If the NNT is not presented, please calculate it.

### Sample NNT for a variety of therapies

Therapy	Event prevented	Length of follow-up	NNT
H. pylori eradication in duodenal ulcer	Ulcer recurrence	1 year	1.1
Finasteride for benign prostatic hypertrophy	Need for surgery	2 years	39
Streptokinase and aspirin for acute myocardial infarction	One death	5 weeks	20
Enalapril for mild/moderate heart failure	One death	1 year	100
Lipid lowering in patients with coronary heart disease	One myocardial infarction (MI) or stroke related death	5 years	16
Treatment of mild high blood pressure	One MI, stroke or death.	1 year	700
Treatment of severe hypertension	One MI, stroke or death	1 year	15

How to calculate:

CER = control event rate

EER = experimental event rate

ARR = CER - EER

NNT = 1 / ARR

What is the NNT for the primary outcome? \_\_\_\_\_

List NNTs for other important secondary outcomes?

\_\_\_\_\_

\_\_\_\_\_

## APPLICABILITY

### Do these results apply to your patient?

Consider

- Is your patient similar to the study population in terms of **important** prognostic variables for treatment success and harms?
- Are costs reasonable?
- Is this treatment available?

Adam C. Ehrlich, MD, MPH

Department of Medicine, Section of Gastroenterology

Temple University School of Medicine

Modified from Deborah Korenstein, MD at Icahn School of Medicine at Mount Sinai



Describe issues of applicability to your patient:

---

---