

# Quantifying treatment effects using randomized trials

## Introduction

---

As we noted in the Preface and Chapter 1, because the purpose of doing diagnostic tests is often to decide whether or how to treat the patient, we may need to quantify the effects of treatment to decide whether to do a test. For example, if the treatment for a disease provides a dramatic benefit, we should have a lower threshold for testing for that disease than if the treatment is of marginal or unknown efficacy. In this chapter, we discuss how to quantify the effects of treatments using the results of randomized trials. In Chapter 10, we will extend the discussion to observational studies of treatment efficacy.

In a randomized trial, study participants are randomly assigned to treatment groups, and the groups are compared to determine which had better outcomes. We begin by briefly reviewing the reasons to do randomized trials, then discuss their critical appraisal. Our approach is somewhat eclectic. Our goal is to highlight issues most important for obtaining and interpreting estimates of treatment effects, not to review the entire topic of randomized trials, and our selection is based partly on issues that seem to have received insufficient attention elsewhere.

We conclude this chapter with a discussion of calculating the treatment cost per bad outcome prevented, a rough step forward in the process of quantifying risks and benefits of treatments.

## Why do randomized trials?

---

The main reason to randomize is to estimate the effect of an intervention without confounding. “Confounding” in this context is the distortion of the estimated treatment effect by extraneous factors associated with the receipt of treatment and causally related to the outcome.<sup>1</sup> This distortion can occur in either direction. Confounding

<sup>1</sup> Terminology for this is not uniform. Some authors refer to this as selection bias. We prefer to refer to it as confounding because many of the methods used to deal with the problem are used to control confounding.

can make a treatment look better than it really is if factors associated with receiving treatment have a favorable effect on outcome. This can happen if, for example, the treatment is more likely to be received by people who are wealthier, better educated, or have better health habits or access to other beneficial treatments. Confounding can make a treatment look worse than it really is if the treatment is more likely to be given to people with a worse prognosis, for example, those who have a particular disease or for whom disease is more severe.

In Chapter 10, we will discuss ways to address the problem of confounding in observational studies of treatments. In this chapter, we discuss randomized trials, which minimize the possibility of confounding as a source of error. Randomization reduces the problem of confounding by creating treatment and control groups likely to be similar with respect to all confounders, both measured and unmeasured, known and unknown. Of course, even with proper randomization, it is possible that the two groups will be different with regard to certain confounders. If the groups do have significant chance asymmetries in important measured confounders, it is possible to control for these by using multivariable analysis or stratification. Most of the time, however, multivariable analyses are not needed.

## **Critical appraisal of randomized trials**

---

Before we turn to quantitatively estimating the effects of treatments, we review some issues in the design, conduct, and analysis of randomized trials that can affect the validity of these estimates.

### **Design and conduct**

#### **Blinding**

“Blinding,” or masking, means keeping the treatment allocation secret. It can be done at three levels: the patient, the care provider, and the person assessing outcome. Blinding the patient prevents differences between groups due to the placebo effect. It is particularly important for subjective outcomes, like pain. Blinding patients in the control group keeps them from finding out that they are not getting active treatment and procuring it outside the study. Blinding the patient as well as the care provider helps avoid differences in co-interventions – that is, other interventions or changes in treatment such as additional care or medications. Finally, blinding the person responsible for outcome ascertainment is important to prevent observer bias. Again, this is most important for subjective outcomes. Thus, blinding the person responsible for outcome ascertainment would not be very important when total mortality is the outcome, but might be important for cause-specific mortality, which, as discussed in Chapter 6, depends on a more subjective process: assigning a cause of death.

### Surrogate outcomes

It is important to distinguish between clinical outcomes the patient can perceive and surrogate outcomes. A “surrogate outcome” is one that is presumed to be associated with the outcome of interest, but is more easily measurable, or occurs more quickly and therefore is more convenient to use in a study. Examples include using changes in levels of risk factors for disease (e.g., blood pressure or bone density) rather than in the development of the disease itself (stroke or fractures), or changes in markers of disease activity or severity (e.g., viral load, hemoglobin A1c) rather than changes in morbidity or mortality from the disease. There are multiple examples of treatments that make the surrogate outcome better but have no effect (or harmful effects) on clinical outcomes of interest (Guyatt et al. 2008). As a general rule, you should be skeptical of studies where the only way the investigators could tell who benefited from an intervention was by doing tests.

### Composite endpoints

In some trials, several possible bad outcomes are grouped together into a composite endpoint. If this composite endpoint combines outcomes of varying importance, it may find a lower risk in the treatment group due entirely to a difference in the risk of a less important outcome. For example, a study (Waksman et al. 2002) of an intervention for blocked coronary artery bypass grafts showed improvement in a composite endpoint of death, myocardial infarction, and revascularization of the target vessel. However, death or myocardial infarction contributed only 5 of the 22 outcome events in the intervention group and 6 of the 43 events in the control group. Thus, there was no evidence that the intervention affected the more important outcomes of death or myocardial infarction – just revascularization. It is even possible for the treatment group to have more of the most important outcomes but sufficiently fewer minor outcomes to mask the increased risk of treatment or even make the composite treatment effect favourable. For example, in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial of aggressive control of blood glucose in adults with diabetes (target haemoglobin A-1c level <6% vs 7–7.9%), the prespecified primary outcome was a composite outcome consisting of nonfatal MI, nonfatal stroke or cardiovascular death. (ACCORD Study Group, 2008). After a mean of 3.5 years of follow up there was a nonsignificant reduction in the risk of the primary outcome. But this negative result masked a statistically significant 1% absolute *increase* ( $P = 0.04$ ) in total mortality that in the composite outcome was balanced by a 1% decrease ( $P = 0.004$ ) in nonfatal myocardial infarction.

This sort of troubling discrepancy between fatal and nonfatal outcomes has been observed in other cardiovascular trials as well. A review (Ferreira-Gonzalez et al. 2007) of 114 randomized trials of cardiovascular interventions that used composite endpoints found that only 68% of the studies reported results for each component of the primary composite endpoint and that outcomes of greater importance to the patient (such as death) were associated with smaller relative treatment effects than less important outcomes.

### Loss to follow-up

Loss to follow-up poses one of the most serious threats to the validity of randomized trials. A good rule to follow is: “once randomized, always analyzed.” However, especially in long-term trials, it is possible to lose track of some study participants and, consequently, not know their outcomes. These losses to follow-up can reduce the power to find a difference simply by reducing the effective sample size, and they can introduce bias in either direction, especially if the reasons for losses to follow-up differ between the treatment groups.

For example, if the patients in the treatment group are lost to follow-up because of some negative effect of the treatment, or the patients in the control group are lost to follow-up because they have recovered from their illnesses, the study will be biased in favor of the treatment. As we described in Chapter 7, a sensitivity analysis can explore the maximum potential bias due to loss to follow-up, with a “worst-case” scenario that assumes poor outcomes for all losses to follow-up in the treatment group and good outcomes for all losses to follow-up in the control group. If a favorable effect of treatment persists, you can be confident that it is not an artifact due to losses to follow-up. More often, this approach will eliminate the treatment benefit or make treatment appear harmful and other approaches will be needed, such as seeking evidence of differences in prognostic factors between subjects lost to follow-up in the two groups.

## Analysis

### Intention-to-treat, as-treated, and per-protocol analyses

When analyzing results in a randomized trial, the groups compared should be based on the treatment assigned rather than the treatment received. This is sometimes called an “intention-to-treat,” as opposed to an “as-treated” analysis, because subjects are analyzed according to the intended treatment. An intention-to-treat analysis is important because patients who complete a course of treatment often have different (usually better) prognoses than patients who do not. For example, in a randomized comparison of coronary artery bypass graft surgery versus medical therapy for stable angina pectoris, some of the patients randomized to surgery became too ill to receive the operation and were treated medically instead (ECSSG 1979). If the results of this trial were analyzed on an “as-treated” basis, those patients too ill to receive surgery would be included in the medical treatment group, which would move patients with the worse prognoses from surgery to medical treatment and bias the results in favor of surgery (Hollis and Campbell 1999).

Between intention-to-treat and as-treated analyses are “per-protocol analyses,” in which only those who were treated according the protocol are analyzed. Although not as obviously biased as an as-treated analysis, a per-protocol analysis is still susceptible to bias, because patients treated according to the protocol are likely to be different from those who are not. A per-protocol analysis in the study of coronary artery bypass graft versus medical therapy would still have excluded patients with worse prognoses from the surgical group (because those patients did not receive the

protocol treatment), but it would not have included them in the medical group. This still would have biased the results in favor of surgery, because the sickest patients would have been removed from that group (Hollis and Campbell 1999).

Dealing with this sort of crossover between groups using intention-to-treat analyses does have a disadvantage. The greater the number of subjects not treated according to the protocol, the less power the study will have, and the more the measure of effect size will be biased toward no effect. Thus, intention-to-treat analysis does not prevent bias; it just ensures that the bias is in a known direction (towards the null).

### Subgroup analyses

The focus of a randomized study is the comparison of the *overall* groups to which subjects are randomized, not comparisons of subgroups. Beware of studies that find no overall difference between treatment and control but highlight a treatment effect in one or another subgroup. If the authors looked at enough subgroups, they were bound to find a treatment effect in one of them. If the overall result is negative and one subgroup did better with treatment, then another subgroup probably did worse.

A classic illustration of the perils of subgroup analysis appeared in the publication of the ISIS-2 (Second International Study of Infarct Survival) results (ISIS-2 1988). This was a randomized trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction. For the purposes of this discussion, the important overall result was that aspirin therapy (1 month of 160 mg/day) reduced cardiovascular mortality from 11.8% to 9.4% ( $P < 0.00001$ ). There is no questioning this overall result in this well-done, randomized, blinded trial. The authors went on to discuss the results of aspirin therapy within several subgroups (diabetics, patients  $\geq 70$  years old, patients with hypertension, a history of prior myocardial infarctions, etc.). They then cautioned readers about these subgroup analyses. To make this point, they divided the study population by astrological sign and showed that Geminis and Libras randomized to aspirin had 11.1% cardiovascular mortality, whereas those randomized to placebo only had 10.2% mortality ( $P = \text{NS}$ ). Of course, given the overall positive effect of aspirin in the study, the patients with the other astrological signs had much lower mortality in the aspirin than in the placebo group (9.0% vs. 12.1%,  $P < 0.00001$ ). Quoting from the paper:

It is, of course, clear that the best estimate of the real size of the treatment effect in each astrological subgroup is given not by the results in that subgroup alone but by the overall results in all subgroups combined.

The ultimate message is to be wary of subgroup analyses, especially those without a strong biological basis.<sup>2</sup>

### Multiple comparisons

Unless there is a breakdown in either the randomization or the blinding, the only way to come up with a falsely positive result in a randomized double-blind trial

<sup>2</sup> We will see in Chapter 11 that the infinitesimally small pre-study probability that astrological sign affects response to aspirin, combined with weak evidence in favor of such an effect, results in a still negligible likelihood of a real effect.

(analyzed according to intention-to-treat) is by chance. The P-value provides only a rough indication of the likelihood of chance as a basis for the association. (We will discuss P-values in Chapter 11.) One of the most common causes for a falsely positive randomized double-blind trial result is that the investigators looked at multiple different outcomes in multiple different groups of patients in multiple different ways.

#### Between-groups versus within-groups comparisons

One would think, having gone to all of the trouble of randomizing the subjects, that investigators would then compare the groups according to group assignment – that is, compare outcomes *between* groups. However, this is not always the case. Sometimes in a randomized trial, investigators will focus on within-group comparisons.

For example, a 2003 study (Nissen et al. 2003) randomized patients with acute coronary syndrome to receive either recombinant ApoA-I Milano or placebo. The authors reported in the abstract that atheroma volume decreased significantly in the treatment group ( $P = 0.02$ ) but not in the control group ( $P = 0.97$ ). However, the P-value for the difference between the two groups (reported in a footnote) was 0.29. Focusing on the within-group changes (in this surrogate outcome) suggested stronger evidence of benefit than the study provided.

#### Direction of biases in randomized blinded trials

If randomization and blinding are done properly, follow-up is reasonably complete, and an intention-to-treat analysis is done, most other problems, such as poor adherence to treatment and random error in the measurement of the outcome variable, will make it harder to find statistically significant differences between the two groups, even if they exist (i.e., the trial will be falsely negative).

The tendency of poorly done studies to be biased toward finding no effect is a particular problem with equivalency trials, where a drug is judged to be effective if it is not demonstrably worse than a drug of known efficacy. In the case of equivalency trials, the normal motivation of investigators to do a trial very carefully in order to maximize the probability of finding a difference between groups is missing. This presents a difficult problem for regulatory agencies. If a treatment is known to be effective, it may not be ethical to randomize people to placebo. But if the investigators' goal for a trial is to demonstrate equivalence, it is easy to do a sloppy job in multiple subtle ways and increase the likelihood of obtaining the desired result.

## Quantifying treatment effects

### Continuous outcome variables

Many randomized trials have continuous outcome variables. For example, in Chapter 3, we estimated the benefit of treatment of influenza with oseltamivir as a reduction in the duration of illness by about one day. It is actually about 32 hours (Treanor et al. 2000). Aside from duration of illness, other examples of good continuous outcome variables are changes in symptom scores, weight, visual acuity, or pain level, because these are not surrogate outcomes but outcomes that the patient can perceive.

**Table 9.1.** Measures of effect size from a randomized trial summarized in a 2 × 2 table

	Bad outcome	No bad outcome	Totals
Treatment	a	b	<b>a + b</b>
Control	c	d	<b>c + d</b>

$$R_T = \text{Risk in Treatment Group} = a/(a + b)$$

$$R_C = \text{Risk in Control Group} = \text{Baseline Risk} = c/(c + d)$$

$$RR = R_T/R_C = \frac{a/(a + b)}{c/(c + d)}$$

$$RRR = 1 - RR$$

$$ARR = -\text{Risk Difference} = -(R_T - R_C) = R_C - R_T = c/(c + d) - a/(a + b)$$

$$NNT = 1/ARR$$

$$OR = ad/bc \text{ (generally should not be used for clinical trials)}$$

The magnitude of differences between groups will depend on the units of measurement. When continuous outcomes are measured on an unfamiliar scale (e.g., a newly created symptom score), they are sometimes standardized by dividing the difference between groups by the standard deviation of the measurement.

**Dichotomous outcome variables**

For dichotomous outcomes, such as death or recurrence of cancer, the treatment effect in a randomized trial can be measured with the risk ratio or relative risk (RR), relative risk reduction (RRR), the absolute risk reduction (ARR), and its inverse, the number needed to treat (NNT). The odds ratio (OR), as discussed below, is overused for measuring treatment effects in randomized trials. These measures are defined in Table 9.1.

A helpful (but by no means universal) convention is to put outcomes in columns and predictors in rows, with the “Bad Outcome” column on the left and the “Treatment” row on the top. When this convention is followed, an  $RR < 1$  means the treatment is beneficial – that is, that it decreases bad outcomes. In contrast, an  $RR > 1$  means the treatment is harmful in some way, as is commonly the case when the bad outcome is a side effect. Box 9.1 gives a specific example, calculating RR, RRR, ARR, and NNT for a randomized blinded trial to prevent breast cancer recurrence in postmenopausal women.

**Relative versus absolute measures of treatment effect**

As was the case in the article cited in Box 9.1, the RRR is the most commonly reported summary measure of treatment effect. To truly understand the effectiveness of the treatment, however, not only relative measures like the RRR and RR, but also absolute measures (ARR and NNT) that account for the baseline risk should be considered. When the baseline risk is low (i.e., the outcome is rare), absolute risks are especially important to report, in order to give a realistic estimate of the likelihood of a treatment benefit for an individual patient. For example, in the letrozole study (Box 9.1), saying that the drug reduces risk of recurrence by 2.2% (from 5.1% to

**Box 9.1: Letrozole and breast cancer recurrence**

A front page headline in the *San Francisco Chronicle* (October 10, 2003) read: “Drug cuts risk of breast-cancer relapse. Findings so promising, study halted so scientists could release news” (Russell 2003). The article reported on a randomized double-blind trial of the aromatase inhibitor letrozole to prevent breast cancer recurrence after tamoxifen therapy in postmenopausal women. For the purposes of this chapter, the key quote from the newspaper story is the following:

The trial was interrupted almost 2½ years after it began. Researchers had scheduled a midpoint peek at the data, and found letrozole was apparently working far better than expected. The women who took it had **43 percent** fewer recurrences of their breast cancer compared to those assigned in the study to take a placebo, or dummy pill.

This letrozole study (Goss et al. 2003) was properly randomized and blinded, used an intention-to-treat analysis, and had minimal losses to follow-up. The trial originally planned to follow subjects for 5 years but was stopped after a median follow-up of 2.4 years. We will use the results to illustrate how to calculate the various measures of treatment effect: RR, RRR, ARR, and NNT:

	Recurrence	No Recurrence	<b>Total</b>
Letrozole	75	2500	<b>2575</b>
Placebo	132	2450	<b>2582</b>

$$\text{Risk(Letrozole)} = 75/2575 = 2.9\%$$

$$\text{Risk(Placebo)} = \text{Baseline Risk} = 132/2582 = 5.1\%$$

$$\text{RR} = \text{Relative Risk or Risk Ratio} = (2.9\%)/(5.1\%) = 0.57$$

$$\text{RRR} = \text{Relative Risk Reduction} = 1 - \text{RR} = 1 - 0.57 = 43\%$$

$$\begin{aligned} \text{ARR} = \text{Absolute Risk Reduction} &= - \text{Risk Difference} \\ &= - (2.9\% - 5.1\%) = 2.2\% \text{ (over 2.4 years)} \end{aligned}$$

**NNT** = Number Needed to Treat =  $1/\text{ARR} = 1/2.2\% = 45$ . This means that we need to treat 45 women with letrozole for an average of 2.4 years to prevent one additional recurrence of breast cancer.

**Treatment Cost per Bad Outcome Prevented:** The dose of letrozole is 2.5 mg/day. The cost of 30 pills (2.5 mg) is about \$266,<sup>3</sup> so the cost of 2.5 years (30 months) of therapy is  $30 \text{ months} \times \$266/\text{month} = \$7980$ . Because we have to treat 45 women to prevent one breast cancer recurrence, the treatment cost per breast cancer recurrence prevented is about  $45 \times \$7980 = \$359,100$ .

2.9%) sounds less impressive than saying that it reduces risk of recurrence by 43%. From the ARR of 2.2%, we can obtain the NNT of 45. What if the absolute recurrence rate in the control group were 1%? Then, the same 43% RRR would translate to an ARR of 0.43% and the NNT would be  $1/0.57\% = 175$ .

<sup>3</sup> (www.drugstore.com, 1/7/08).

### Inflating the apparent effect size by using the odds ratio

The OR (Table 9.1) is another measure of treatment effect that is sometimes reported. However, it is generally not necessary or desirable to report the OR as a measure of effect size in a randomized controlled trial. The OR is an appropriate measure of association for case–control studies and a natural output of observational studies that use logistic regression to control for confounding. However, the RR has a much more natural and intuitive interpretation than the OR. Perhaps the reason that investigators sometimes use OR to report treatment effects in randomized controlled trials is that the OR is always farther from 1 than the RR. This can make results seem much more impressive than they are, especially when the outcome is relatively common. For example, in a randomized trial of varenicline to support smoking cessation, the 13- to 24-week abstinence rate was 70.5% with varenicline, compared with 49.6% with placebo (Tonstad et al. 2006). The authors reported the OR of 2.48, which is more impressive than the RR of 1.42. (They also did not follow the convention of putting the bad outcome, resumption of smoking, on the top.)

### Treatment cost per bad outcome prevented

In Chapter 1, we presented the following clinical scenario.

#### Clinical Scenario: Flu Prophylaxis

It is the flu season, and your patient is a 14-year-old girl with fever, myalgias, cough, and sore throat persisting for 1 day. Her mother has seen a commercial for Tamiflu® (oseltamivir) and asks you about prescribing it for the whole family so they don't catch the flu. Other family members in the household are the child's parents, her two younger brothers, and her maternal grandparents (in their 70s). They are currently well, but none has had the flu shot this year.

You are considering testing the patient with a rapid bedside test for influenza A and B.

Welliver et al. (2001) addressed the issue of prophylactic oseltamivir in household contacts of patients with the flu. Their study was a randomized blinded trial of oseltamivir (Tamiflu®; 75 mg/day for 5 days) to prevent influenza in the household contacts of patients with flu-like symptoms during the 1998–1999 flu season. The results were stratified by whether the index case had laboratory-proven influenza (415 subjects) or not (540 subjects). This study was properly randomized and blinded, used an intention-to-treat analysis, and had minimal losses to follow-up.

The results of oseltamivir prophylaxis in household contacts of index cases with laboratory-proven influenza are shown in Table 9.2.

When the index case had laboratory-proven influenza, the baseline risk of the family contacts getting symptomatic influenza in the placebo group was 12.6%. The oseltamivir prophylaxis reduced this risk to 1.4%, a RRR of 89%.

The results of prophylaxis when the index cases did not have influenza suggested a nearly identical RRR, but a much lower baseline risk of getting symptomatic influenza (Table 9.3). In these family contacts of a flu-negative index case, the baseline risk of influenza was only 3.1%. The prophylaxis reduced this risk to a risk of 0.4%, again a RRR of 89%.

**Table 9.2.** Results of oseltamivir prophylaxis in household contacts of patients with laboratory proven influenza

Index case flu+	Household contacts			Risk
	Flu	No flu	Total	
Oseltamivir	3	206	<b>209</b>	3/209 = 1.4%
Placebo	26	180	<b>206</b>	26/206 = 12.6%
<b>Total</b>	<b>29</b>	<b>386</b>	<b>415</b>	

RR: 1.4%/12.6% = 0.114

RRR: 1 – RR = 89%

ARR: 12.6% – 1.4% = 11.2%

NNT<sup>4</sup>: 1/ARR = 9

If the RRRs were reported without the baseline risks, we would have no way of knowing how much better it is to treat a household contact when the index case is positive than when the index case is negative; the RRR was 89% in both groups.

In the letrozole example in Box 9.1, we introduced the idea of multiplying the NNT by the treatment cost ( $C_{\text{treat}}$ ) to get the treatment cost of preventing one bad outcome:

NNT = Number Needed to Treat to Prevent One Bad Outcome

$C_{\text{treat}}$  = Cost of One Treatment

$\text{NNT} \times C_{\text{treat}}$  = Treatment Cost/Bad Outcome Prevented

The ratio of treatment costs to bad outcomes prevented is a “cost effectiveness ratio.” It is not *the* cost effectiveness ratio, because there are other ways to calculate a ratio of costs to outcomes prevented (see Box 9.2.) Tamiflu<sup>®</sup> costs about \$85 for ten

**Table 9.3.** Results of oseltamivir prophylaxis in household contacts of index cases who did not have influenza

Index case flu–	Household contacts			Risk
	Flu	No flu	Total	
Oseltamivir	1	283	<b>284</b>	1/284 = 0.4%
Placebo	8	248	<b>256</b>	8/256 = 3.1%
<b>Total</b>	<b>9</b>	<b>531</b>	<b>540</b>	

RR: 0.4%/3.1% = 0.113

RRR: 1 – RR = 89%

ARR: 3.1% – 0.4% = 2.8%

NNT: 1/ARR = 36

<sup>4</sup> Strictly speaking, when the outcome prevented is an infectious disease, the number needed to treat should be decreased to account for the decreased transmission beyond the first-level contact. If  $\alpha$  is the probability that one individual will pass the disease to another, the “discount factor” is  $(1 - \alpha)$ . We ignore the issue of second- and higher-generation infections because we want this discussion to apply to bad outcomes other than infectious diseases.

**Box 9.2: Treatment cost per bad outcome prevented vs. traditional cost effectiveness analysis and cost benefit analysis**

Multiplying the NNT by the treatment cost does not consider the dollar costs associated with the bad outcome that is prevented by the treatment. Thus, what we are calculating is the treatment cost of preventing one bad outcome along with all dollar costs associated with that bad outcome:

$C_{\text{treat}}$  = Cost of One Treatment

$C_{\text{outcome}}$  = Cost of One Bad Outcome

$$\text{NNT} \times C_{\text{treat}} = \frac{\text{Treatment Cost}}{\text{Bad Outcome Prevented (including its costs)}}$$

Because it is often much easier to estimate treatment cost ( $C_{\text{treat}}$ ) than the costs associated with a bad outcome ( $C_{\text{outcome}}$ ), we prefer to leave these outcome costs in the denominator. If, however, we are willing to estimate the bad outcome's costs, we can subtract these costs from the numerator and more closely approximate the cost effectiveness ratio that would result from a traditional cost effectiveness analysis.

$$(\text{NNT} \times C_{\text{treat}}) - C_{\text{outcome}} = \frac{\text{Treatment Cost} - C_{\text{outcome}}}{\text{Bad Outcome Prevented}}$$

In our example of flu prophylaxis, suppose the dollar costs of a case of the flu amount to \$400. (This would include all of the dollar costs we can easily count,<sup>5</sup> including missed work, ibuprofen for myalgias and fever, some chance of a doctor visit, etc., but not the discomfort of having the flu.) Our original treatment cost per flu case prevented was \$405. Now, because each flu case is associated with \$400 in costs, the new cost effectiveness ratio is  $\$405 - \$400 = \$5$  per flu case prevented. Considering the costs of a bad outcome along with treatment costs to prevent the bad outcome leads to a lower total cost per bad outcome prevented. For the rest of this chapter, we will lump the costs associated with a bad outcome along with the bad outcome itself in the denominator.

Once we account for all of the financial costs, we are left with the cost to prevent the discomfort of the flu. Theoretically, we could assign a dollar value to that, too. Then we could do a cost-benefit analysis, in which we simply add up all of the costs and benefits and see if we come out ahead. For example, if we assign a dollar value of \$440 for the discomfort of having the flu, then we could say we spend a net \$5 on medication and get \$440 worth of relief of suffering in return, so we would come out \$435 ahead for each case of the flu we prevent. (Of course, we might do even better with the influenza vaccine!)

75-mg pills.<sup>6</sup> We assume that a prophylactic course (5 pills) would cost about \$45. With this treatment cost ( $C_{\text{treat}}$ ), we can calculate the cost of preventing a case of influenza if the index case is influenza-positive (Flu+) or influenza-negative (Flu-).

Index Case Flu+:

$\text{NNT} = 9$  (Treat 9 household contacts, prevent 1 flu case.)

$\text{NNT} \times C_{\text{treat}} = 9 \times \$45 = \$405/\text{flu case prevented}$

<sup>5</sup> Something that we can't easily count, but probably should be considered, is the possibility of contributing to the development of resistance to oseltamivir.

<sup>6</sup> [www.drugstore.com](http://www.drugstore.com) 1/7/08.

Index Case Flu–:

$NNT = 36$  (Treat 36 household contacts, prevent 1 flu case.)

$NNT \times C_{\text{treat}} = 36 \times \$45 = \$1620/\text{flu case prevented}$

The RRR associated with treating the contacts of Flu– index cases is the same as for contacts of Flu+ index cases. However, the baseline risk of contracting influenza is four times lower, so the absolute benefit is four times lower, and the cost per flu case prevented is four times higher. In fact, for the rest of this discussion, we will assume that the benefit of treating the household contacts of Flu– patients is negligible. In Chapter 1, we discussed the definition of disease and the assumption that nondiseased patients would not benefit from treatment. In this case, the “disease” is being the household contact of a Flu+ index case. While the “nondiseased” contacts of a Flu– case would, in fact, benefit slightly, this benefit can be ignored with minimal loss of accuracy.

#### Uncertainty about whether the patient has the disease

When we do not know whether our patient has the disease (D+) and will benefit from treatment, we must adjust our NNT upward (again assuming no benefit to treating D– patients). The adjusted NNT (NNT\*) is the original number needed to treat (calculated only using D+ patients) divided by the probability (P) of disease.

$NNT = \text{Number Needed to Treat to prevent one bad outcome assuming that all those treated are D+}$

$P = \text{Probability of D+}$

$NNT^* = \text{Adjusted Number Needed to Treat} = NNT/P$

In our flu prophylaxis example, we do not know if our 14-year-old patient actually has influenza or some other viral illness. The prevalence of laboratory-proven influenza in the oseltamivir prophylaxis study (Welliver et al. 2001) was about 45%. Assuming that this is the prevalence of influenza in our population, the expected ARR is only 45% as large, so the NNT will be 1/0.45 times as high<sup>7</sup>:

$NNT \text{ if index case Flu+} = 9.$

$\text{Probability of flu} = P = 0.45$

$ARR^* = 0.45 \times ARR$

$NNT^* = NNT/P = 9/0.45 = 20$

$NNT^* \times C_{\text{treat}} = \text{Treatment Cost/Bad Outcome Prevented}$

$\text{Cost/Case Prevented} = 20 \times \$45 = \$900$

Now suppose we decide that it is worth no more than \$1080 in treatment costs to prevent a case of the flu (and its associated costs). Because each patient we treat costs

<sup>7</sup> More generally, the ARR will be a weighted average of the ARR in those in whom the index case does and does not have the flu. In this case, it would be  $ARR = P \times 11.2\% + (1 - P) \times 2.8\%$

\$45, we can say that if the  $NNT^*$  is more than  $\$1080/\$45 = 24$ , then it would not be worth treating. So now the question is: at what probability of influenza in the index case  $P_{TT}$  will the  $NNT^*$  be no more than 24? Again, neglecting the small benefit from treating influenza-negative index patients, that probability will be when  $NNT^* = NNT/P_{TT}$  (see above), so it will be when  $P_{TT} = NNT/NNT^* = 9/24 = 0.375$ . Thus, if you treat household contacts when the index case’s probability of the flu is 37.5% or higher, you will not spend more than \$1080 per case of flu prevented. Now, assume that you can test for the flu. In Chapter 3, we discussed how to use  $P_{TT}$  and the test characteristics  $[LR(+)$  and  $LR(-)]$  to calculate lower and upper probabilities where a testing strategy could make sense.<sup>8</sup>

**“Number needed to harm”**

To this point we have only considered the trade-off between the dollar costs of treatment and the effectiveness of treatment in preventing bad outcomes. Treatments are often associated with undesired effects that should also be balanced against the reduced risk of the primary outcome. Undesired effects of treatment can be evaluated using the same kind of  $2 \times 2$  table as desired effects. In our flu prophylaxis example, oseltamivir was more frequently associated with nausea than was placebo (Table 9.4).

The RR for a side effect like nausea is greater than one ( $RR = 2.1$ ), and because the risk of the bad outcome is higher in the treatment group than in the control group, the ARR is negative. Because we prefer dealing with positive numbers, we calculate an absolute risk increase ( $ARI = -ARR$ ), rather than an ARR. The number needed to harm (NNH)<sup>9</sup> is defined as  $1/ARI$ , so it is actually the number of patients treated for each one harmed. In this case, the NNH is 35, so for every 35 patients treated, we will cause one additional case of nausea.

In some cases, especially when a treatment is associated with severe or common side effects, we might be interested in quantifying the trade-off between side effects and primary outcome prevention, rather than the trade-off between dollar costs

**Table 9.4.** Association of oseltamivir prophylaxis with nausea

	Household contacts		Total	Risk
	Nausea	No nausea		
Oseltamivir	27	467	494	$27/494 = 5.5\%$
Placebo	12	449	461	$12/461 = 2.6\%$

RR:  $5.5\%/2.6\% = 2.10$   
 ARR:  $2.6\% - 5.5\% = -2.9\%$   
 ARI:  $5.5\% - 2.6\% = 2.9\%$   
 NNH:  $1/ARI = 35$

<sup>8</sup> In terms of Chapter 3, net benefit B of treating a household contact of a flu+ individual is the expected benefit of treatment less the cost of treatment:  $1/9 \times \$1080 - \$45 = \$75$ . The cost C of treating a household contact of a flu- individual (who we assume will not benefit) is \$45. The threshold odds are  $C/B = \$45/\$75 = 0.6$ , and the threshold probability is  $0.6/(1 + 0.6) = 0.375$ . Since you would use the test on the index case to guide treatment on 6 individuals, the testing cost T is the cost of testing the index case divided by 6.

<sup>9</sup> “Number Needed to Harm” is an established term that really means “Number Needed to Treat to Cause Harm in One.”

and outcome prevention. For example, we might be more interested in the trade-off between the nausea associated with oseltamivir treatment and the reduced likelihood of the flu. We can then calculate the number of “harms” per primary outcome prevented. This is simply the ARI for the undesired effect divided by the ARR for the primary outcome, or equivalently, the NNT divided by the NNH:

“Harms”/Bad Outcome Prevented = ARI/ARR = NNT/NNH.

In the case of flu prophylaxis,

Cases of Nausea/Flu Case Prevented = ARI/ARR = NNT/NNH = 9/35 ~ 1/4

This means that we cause one-fourth of a case of nausea for each case of the flu that we prevent, or one case of nausea caused for every four cases of the flu prevented. This is an attractive trade-off because a case of the flu is typically much worse than an episode of medication-induced nausea. The ratio of side effects caused per primary outcome prevented is a cost effectiveness ratio in which the costs are measured in side effects caused instead of dollars spent.

For the nausea associated with oseltamivir prophylaxis and for most other side effects of treatment, the NNH should not be adjusted for the likelihood of disease, as there is no reason to believe that oseltamivir is less likely to cause nausea in household contacts of Flu– patients than in those of Flu+ patients.

## Summary of key points

1. In a randomized blinded trial of a treatment, the purpose of the randomization is to ensure that, at baseline, the groups are similar with respect to confounders, both known and unknown.
2. The purpose of the blinding is to prevent differential co-interventions and biased outcome assessment.
3. In order to preserve the value of randomization, the study should compare the randomized groups in an intention-to-treat analysis and minimize losses to follow-up.
4. Use caution with studies using surrogate outcomes or relying on subgroup analysis to show a treatment effect.
5. When the outcome of a randomized blinded trial is dichotomous, such as death or recurrence of cancer, one assesses the treatment effect by comparing the outcome risk in the treatment and the control groups. The ratio of these risks is the risk ratio; the difference between them is the absolute risk reduction.
6. The inverse of the absolute risk reduction is the number needed to treat to prevent one outcome.
7. The treatment cost per bad outcome prevented is simply the number needed to treat times the cost of treatment.
8. In the case of side effects, the risk of the undesired outcome is higher in the treatment than the control group; the risk difference is the absolute risk increase, and its inverse is the number needed to harm.
9. Using the results of randomized blinded trials to balance the effectiveness of treatment in preventing undesired outcomes with the costs and risks of

unnecessary treatment helps determine the treatment threshold – that is, the likelihood of disease at which treatment is indicated.

10. This treatment threshold probability can, in turn, be used with test characteristics to determine when a diagnostic test might be helpful in guiding treatment decisions.

## References

- Accord Study Group (2008). “Effects of intensive glucose lowering in type 2 diabetes.” *N Engl J Med* **358**(24): 2630–3.
- ECSSG (1979). “Coronary-artery bypass surgery in stable angina pectoris: Survival at two years. European Coronary Surgery Study Group.” *Lancet* **1**(8122): 889–93.
- Ferreira-Gonzalez, I., J. W. Busse, et al. (2007). “Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials.” *Br Med J* **334**(7597): 786.
- Goss, P. E., J. N. Ingle, et al. (2003). “A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer.” *N Engl J Med* **349**(19): 1793–802.
- Guyatt, G., D. Rennie, et al. (2002). *Users’ guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago, IL, AMA Press.
- Hollis, S., and F. Campbell (1999). “What is meant by intention to treat analysis? Survey of published randomised controlled trials.” *Br Med J* **319**(7211): 670–4.
- Hulley, S. B., S. R. Cummings, et al. (2007). *Designing Clinical Research: An Epidemiologic Approach*. 3<sup>rd</sup> Ed. Philadelphia, PA, Lippincott Williams & Wilkins.
- ISIS-2 (1988). “Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group.” *Lancet* **2**(8607): 349–60.
- Nissen, S. E., T. Tsunoda, et al. (2003). “Effect of recombinant ApoA-I Milano on coronary atherosclerosis in patients with acute coronary syndromes: a randomized controlled trial.” *JAMA* **290**(17): 2292–300.
- Russell, S. (2003). Drug cuts risk of breast-cancer relapse. *Chronicle*. San Francisco, CA. Oct 10, Page 1.
- Tonstad, S., P. Tonnesen, et al. (2006). “Effect of maintenance therapy with varenicline on smoking cessation: a randomized controlled trial.” *JAMA* **296**(1): 64–71.
- Treanor, J. J., F. G. Hayden, et al. (2000). “Efficacy and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza: a randomized controlled trial. US Oral Neuraminidase Study Group.” *JAMA* **283**(8): 1016–24.
- Waksman, R., A. E. Ajani, et al. (2002). “Intravascular gamma radiation for in-stent restenosis in saphenous-vein bypass grafts.” *N Engl J Med* **346**(16): 1194–9.
- Welliver, R., A. S. Monto, et al. (2001). “Effectiveness of oseltamivir in preventing influenza in household contacts: a randomized controlled trial.” *JAMA* **285**(6): 748–54.

## Chapter 9 Problems: quantifying the treatment effects

1. Otitis media with effusion (OME) is common in infants and young children. The basic problem is that the eustachian tube does not work well, and the kids get fluid and negative pressure in the middle ear, which can cause mild to moderate conductive hearing loss and an increased risk of acute (purulent) otitis media (ear infection).

A controversial clinical trial (Mandel et al. 1987) found that, in children who had OME for 3 months, resolution rates at 4 weeks were about 30% with the antibiotic amoxicillin (with or without an antihistamine/decongestant) and about 14% with placebo.

- a) Using the conventions suggested in the chapter (RR is the risk of something bad in the treatment group relative to the control group), what are the RR, RRR, ARR, and NNT to prevent one persistent effusion?
- b) Why are the RRR and ARR so similar in this case?

The reason why the study was so controversial is that one of the investigators (Erdem Cantekin) so disagreed with the other investigators that he published an alternative report on the same study in *JAMA* (Cantekin et al. 1990, 1991; Rennie 1991) after the other investigators reported the results in the *New England Journal of Medicine*. One of Cantekin's main points was that blinding was suspect and no benefit was apparent when the outcome was assessed objectively (by tympanometry). After excluding 43 children (13.3% of the placebo group and 7.4% of amoxicillin group;  $P = 0.122$ ) who had developed acute otitis media during the follow-up period, he came up with the following numbers (simplified from his Table 3):

Outcome Measure	Amoxicillin (%)	Placebo (%)	Difference (%)	P
Normal by otoscopy	35.2	19.2	16.0	0.004
Normal by algorithm (defined in protocol)	25.6	13.9	11.7	0.027
Normal by tympanometry	17.8	10.0	7.8	0.121
Normal by hearing test	21.9	18.0	3.9	0.611
Hearing improved >10 dB	31.5	32.5	-1.0	0.311

- c) Do you agree with the decision to exclude children who developed acute otitis during the follow-up period? What effect might this have had on the results tabulated above?
  - d) Do you agree with Cantekin's conclusion that amoxicillin was not effective, based on the results for the most objective outcome measurement (tympanometry)?
2. Children with bacterial meningitis can suffer brain damage or hearing loss, referred to respectively as neurologic or audiologic sequelae. Odio et al. (1991) did a randomized blinded trial of the corticosteroid dexamethasone to prevent neurologic or audiologic sequelae in 99 infants and children with bacterial meningitis. From the "Results" section of their abstract:

At follow-up examination, 7 of the . . . 51 dexamethasone-treated children (14 percent) and 18 of 48 . . . controls (38 percent) had one or more neurologic or audiologic sequelae ( $P = 0.007$ ); the relative risk of sequelae for a child receiving placebo as compared with a child receiving dexamethasone was 3.8 (95 percent confidence interval, 1.3 to 11.5) . . .

- a) The risk of sequelae was lower in the dexamethasone group (14%) than in the control group (38%). In other words, the dexamethasone was beneficial. Why is the reported RR greater than 1?
  - b) Even allowing for part (a), the reported RR is incorrect. Calculate the correct RR.
  - c) How did the authors get the wrong number?
3. Patients with chronic hepatitis who have the “e” antigen of the hepatitis B virus in their blood stream (HBeAg), tend to suffer worse liver damage. Patients with chronic hepatitis are also commonly treated with interferon alfa. Read the following abstract with this question in mind: Does treatment with interferon alfa improve clinical outcome in this group of patients?”

Long-Term Follow-Up of HBeAg-Positive Patients Treated with Interferon Alfa for Chronic Hepatitis B (Niederau et al. 1996)

**Background** In patients with chronic hepatitis B, treatment with interferon alfa and the consequent loss of hepatitis B e antigen (HBeAg) from the blood leads to a reduction in inflammatory activity, but the clinical benefits of this treatment have not been established. **We evaluated whether HBeAg seroconversion induced by interferon alfa improves clinical outcome.** [Emphasis added]

**Methods** We studied prospectively a cohort of 103 patients treated with interferon alfa for chronic hepatitis B; the mean ( $\pm$ SD) follow-up was  $50.0 \pm 19.8$  months. Fifty-three untreated patients served as controls.

**Results** After treatment with interferon alfa, 53 of 103 patients no longer had detectable HBeAg or hepatitis B virus DNA, although only 10 patients became seronegative for hepatitis B surface antigen (HBsAg) . . . Of the 53 untreated patients, only 7 spontaneously eliminated HBeAg . . . and all remained positive for HBsAg ( $P < 0.001$  for the comparison with the treated patients, by the proportional-hazards model). During follow-up, 6 of the 103 treated patients died of liver failure, and 2 needed liver transplantation; all 8 were persistently positive for HBeAg. In another eight treated patients, complications of cirrhosis developed; all but one of these patients remained positive for HBeAg. Overall survival and survival without clinical complications were significantly longer in patients who were seronegative for HBeAg after therapy with interferon alfa than in those who remained seropositive ( $P = 0.004$  and  $P = 0.018$ , respectively). In a regression analysis, clearance of HBeAg was the strongest predictor of survival. Of the 53 untreated patients, 13 had severe complications (including 4 deaths and 1 need for liver transplantation); all 13 continued to be HBeAg-positive.

**Conclusions** In patients with chronic hepatitis B infection, the clearance of HBeAg after treatment with interferon alfa is associated with improved clinical outcomes.

- a) The treatment group had 103 patients, and the control group had 53 patients. How were patients assigned to these groups?
  - b) Create a  $2 \times 2$  table comparing the composite outcome of liver transplant or death from liver failure in the treated and untreated patients. Lay out the  $2 \times 2$  table according to the conventions suggested in the chapter and calculate the risk (of the composite outcome) in the control group,  $R_C$ , and the risk in the treatment group,  $R_T$ . Also calculate the RR.
  - c) Is the conclusion of the abstract correct? Discuss.
4. An abstract of a trial of intranasal lidocaine (a local anesthetic) for treatment of migraine headaches (Maizels 1996) is excerpted below:

**OBJECTIVE:** To evaluate the effectiveness of intranasal lidocaine for treatment of acute migraine headache.

**DESIGN:** Prospective, randomized, double-blind, placebo-controlled trial.

**SETTING:** Community urgent care department.

**PATIENTS:** A total of 81 patients (67 women and 14 men; median age, 42 years; range, 19–68 years) with a chief complaint of headache who fulfilled criteria of the International Headache Society for migraine participated . . .

**INTERVENTION:** Patients were randomized in a 2:1 ratio to receive a 4% solution of intranasal lidocaine or saline placebo, respectively.

**MAIN OUTCOME MEASURES:** The primary outcome measure was at least 50% reduction of headache within 15 minutes after treatment . . .

**RESULTS:** Of 53 patients who received intranasal lidocaine 29 (55%) had at least a 50% reduction of headache compared with 6 (21%) of 28 controls ( $P = .004$ )

**CONCLUSIONS:** Intranasal lidocaine provides rapid relief of headache in approximately 55% of ambulatory patients with migraine . . .

- a) Is the main outcome variable in this study subjective or objective?
  - b) Intranasal administration of 4% lidocaine might feel different to the patient than administration of saline. What problem could this cause?
  - c) Do you agree with the conclusion that: “Intranasal lidocaine provides rapid relief of headache in approximately 55% of ambulatory patients with migraine?” Why or why not?
5. In the chapter, we learned from the ISIS-2 Study (ISIS-2 1988) that aspirin therapy (1 month of 160 mg/day) in patients with acute myocardial infarction (AMI) reduced 30-day cardiovascular mortality from 11.8% in the placebo group to 9.4% in the aspirin group ( $P < 0.00001$ ).
- a) What is the ARR?
  - b) How many AMI patients need to be treated with aspirin for 30 days to prevent one death?

- c) A 120-pill bottle of 81-mg aspirin tablets costs about \$5.00. Considering only the cost of the treatment drug, use this to calculate the cost per bad outcome prevented for aspirin therapy of AMI. (Be careful, you only need 60 aspirin pills to treat 1 patient for 30 days.)

ISIS-2 and other studies also showed that, when added to aspirin, thrombolysis with streptokinase reduced mortality even further. The GUSTO Study (GUSTO 1993) compared thrombolysis using tissue plasminogen activator (tPA) with thrombolysis using streptokinase (SK) in reducing the 30-day mortality of patients with AMI receiving aspirin. This study showed a 14% RRR in the tPA-treated patients. Controversy exists because other large studies found no difference in 30-day mortality. Also, only the subgroup of American (not European) patients showed a benefit, and it was not a blinded study. However, for this problem, assume that the 14% RRR is real. The 30-day mortality risk in the SK group was 7.3%.

- d) What was the 30-day mortality risk in the tPA group?
- e) What was the ARR associated with tPA (vs. SK)?
- f) How many patients need to be treated with tPA instead of SK to prevent one death?
- g) The cost of tPA is about \$3,400 per course of treatment, and the cost of SK is about \$560 (Peacock et al. 2007). What is the approximate treatment drug cost per death prevented by using tPA instead of SK?
6. A randomized, double-blind, multi-center trial of tegaserod (Zelnorm<sup>®</sup>) for chronic constipation in 1,264 patients (Kamm et al. 2005) reported response rates of 40.2% with tegaserod 6 mg twice a day, compared with 26.7% with placebo ( $P = 0.0059$ ). (A 2-mg dose was less effective.) The primary outcome was an increase of at least 1 complete spontaneous bowel movement (CSBM) per week over a 4-week period, compared with a 2-week baseline period.
- a) If we consider those who did not respond to be treatment failures, what are the RR, RRR, and ARR (for treatment failure) for tegaserod 6 mg twice a day in this study?
- b) What is the approximate NNT (with 6 mg Zelnorm twice a day for 4 weeks) for each patient that responds by having at least one CSBM per week?
- c) Before it was taken off the market for causing heart attacks, Zelnorm<sup>®</sup> cost about \$180 for 60 pills (<http://www.drugstore.com>, accessed 11/9/06). What is the cost of a week of treatment at the dose above per patient that responds?
- d) The endpoint of this study was an increase of  $\geq 1$  CSBM. If each of the responders actually has two additional CSBM per week, what is the approximate cost of Zelnorm<sup>®</sup> per CSBM?

## References for problem set

- Cantekin, E. I., T. W. McGuire, et al. (1990). "Biomedical information, peer review, and conflict of interest as they influence public health." *JAMA* 263(10): 1427–30.

- Cantekin, E. I., T. W. McGuire, et al. (1991). "Antimicrobial therapy for otitis media with effusion ('secretory' otitis media)." *JAMA* **266**(23): 3309–17.
- GUSTO (1993). "An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The GUSTO investigators." *N Engl J Med* **329**(10): 673–82.
- ISIS-2 (1988). "Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group." *Lancet* **2**(8607): 349–60.
- Kamm, M. A., S. Muller-Lissner, et al. (2005). "Tegaserod for the treatment of chronic constipation: a randomized, double-blind, placebo-controlled multinational study." *Am J Gastroenterol* **100**(2): 362–72.
- Maizels, M., B. Scott, et al. (1996). "Intranasal lidocaine for treatment of migraine: a randomized, double-blind, controlled trial." *JAMA* **276**(4): 319–21.
- Mandel, E. M., H. E. Rockette, et al. (1987). "Efficacy of amoxicillin with and without decongestant-antihistamine for otitis media with effusion in children. Results of a double-blind, randomized trial." *N Engl J Med* **316**(8): 432–7.
- Niederau, C., T. Heintges, et al. (1996). "Long-term follow-up of HBeAg-positive patients treated with interferon alfa for chronic hepatitis B." *N Engl J Med* **334**(22): 1422–7.
- Odio, C. M., I. Faingezicht, et al. (1991). "The beneficial effects of early dexamethasone administration in infants and children with bacterial meningitis." *N Engl J Med* **324**(22): 1525–31.
- Peacock, W. F., J. E. Hollander, et al. (2007). "Reperfusion strategies in the emergency treatment of ST-segment elevation myocardial infarction." *Am J Emerg Med* **25**(3): 353–66.
- Rennie, D. (1991). "The Cantekin affair." *JAMA* **266**(23): 3333–7.