Methodology and Application of Clinical Trials in Radiology: A Self-Assessment Module

William C. Black, MD
Department of Radiology
Dartmouth-Hitchcock Medical Center

Elizabeth A. Krupinski, PhD
Department of Radiology Research
University of Arizona

Annemarie Relyea-Chew, MD, MS
Department of Radiology
University of Washington

Felix S. K. Chew, MD, MBA
Department of Radiology
University of Washington

AJR2008:190:S23-S28, American Roentgen Ray Society

Methodology and Application of Clinical Trials in Radiology: Self-Assessment Module

ABSTRACT

The educational objectives of this self-assessment module are for the participant to understand the differences between cohort and case-control studies in radiology; to understand the advantages of randomized controlled trials over observational studies; to understand the basic principles underlying the use of imaging examinations for screening asymptomatic populations for particular diseases; and to understand the biases associated with the use of survival statistics in the evaluation of screening.

INTRODUCTION

This self-assessment module on clinical trial methodology has an educational component and a self-assessment component. The education component consists of four required articles that the participant should read. The self-assessment component consists of ten multiple-choice questions with solutions. All of these materials are available on the ARRS Web site (www.ARRS.org).

EDUCATIONAL OBJECTIVES

By completing this module, the participant will:

1. Understand the differences between cohort and case-control studies in radiology;
2. Understand the advantages of randomized controlled trials over observational studies;
3. Understand the basic principles underlying the use of imaging examinations for screening asymptomatic populations for particular diseases; and
4. Understand the biases associated with the use of survival statistics in the evaluation of screening.

REQUIRED READING

1. Black WC, Welch HG. *Screening for disease.* AJR 1997; 168(1): 3-11.
2. Blackmore CC. *The challenge of clinical radiology research.* AJR 2001; 176(2):327-31.
3. Blackmore CC, Cummings P. *Observational studies in radiology.* AJR 2004; 183(5):1203-1208.
4. Stolberg HO, Norman G, Trop I, *Randomized controlled trials.* AJR 2004.; 183(6): 1539-1544.

RECOMMENDED READING

1. Fletcher RD, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials. Third edition. Philadelphia: Lippincott Williams & Wilkins, 1996

2. Hunink MGM, Glasziou PP, et al. Decision making in health and medicine: integrating evidence and values. Cambridge, UK: Cambridge University Press, 2001

INSTRUCTIONS

1. Complete the required reading.
2. Visit www.arrs.org and go to the left-hand menu bar under Publications/Journals/SAM articles.
3. Using your member login, order the online SAM as directed.
4. Follow the online instructions for entering your responses to the self-assessment questions and complete the test by answering the questions online.

QUESTIONS

Question 1
Other things being equal, evidence-based medicine (EBM) has established a ranking of levels of evidence based on study design. Which one of the following statements about the relative strengths of these levels of evidence is true?

A. Case control studies are considered stronger than cohort studies.
B. Expert opinion is considered stronger than case control studies.
C. Case series are considered stronger than case control studies.
D. Randomized controlled trials are considered stronger than cohort studies.
E. Expert opinion is considered stronger than a case series.

Question 2
Which of the following is an advantage of a cohort study design over a case-control study design?

A. Requires less time to complete.
B. Eliminates bias in assessment of outcomes.
C. Does not require IRB approval.
D. Outcomes are assessed prospectively.
E. Many different exposures can be studied.

Question 3
Which of the following is an advantage of a case-control study design over a cohort study design?

A. More efficient for investigating rare outcomes.
B. Not subject to recall bias.
C. Does not require IRB approval.
D. Exposures can be assessed prospectively.
E. Possibility of confounding is eliminated.

Question 4
Confounding can occur when two groups being compared differ with respect to some factor associated with the outcome. Which one of the following approaches best minimizes the effect of all types of potential confounding?

A. Stratifying the subjects on the basis of potential confounders.
B. Adjusting for potential confounders with regression analysis.
C. Matching subjects on the basis of potential confounders.
D. Careful allocation of subjects by experienced investigators.
E. Randomly assigning subjects to each of the groups.

Question 5
In a hypothetical case-control study of smoking and lung cancer, 105 cases of newly diagnosed lung cancer were identified in the admission records of a large hospital over a defined period of time. Over the same period of time, 300 age-matched patients admitted to the hospital for reasons other than lung cancer were

selected as controls. Based on a retrospective review of the medical records, 100 of the 105 cases and 200 of the 300 controls were cigarette smokers (see Table 3).

Table 3. Data from Hypothetical Case-Control Study

	Lung Cancer Cases	Non-Lung Cancer Controls
Smokers	100	200
Non Smokers	5	100

Based on these hypothetical results, which conclusion can be drawn?

A. The prevalence of smoking in the general population is about 75%.
B. Smoking causes lung cancer.
C. About 33% of smokers will develop lung cancer.
D. Smokers are ten times more likely to develop lung cancer.
E. About 5% of nonsmokers will develop lung cancer.

Question 6
Percent Five-year survival from diagnosis is a statistic that is often used in studies involving diseases with significant mortality. In the context of a hypothetical study of a new screening program, investigators compared five-year survival between the screen detected cases in the program and cases presenting with symptoms outside the program. Which one of the following statements about these comparisons of five-year survival is true?

A. They are not affected by lead time bias.
B. They are affected by recall bias.
C. They are not affected by selection bias.
D. They are affected by overdiagnosis bias.
E. They are not affected by length time bias.

Question 7
Which one of the following statements about randomized controlled trials of screening is true?

A. Ascertainment bias in determining the cause of death does not affect the estimation of the screening effect.
B. Lack of compliance in the screened group results in an overestimation of the screening effect.
C. Lack of compliance in the control group results in an overestimation of the screening effect.
D. The major purpose of randomization is to minimize the differences between subjects in the screened group and the control group.
E. Subjects who do not comply with their randomized assignment should not be included in the analysis of the screening effect.

Question 8
Which one of the following is a consequence of overdiagnosis in screening?

A. A decrease in the calculated sensitivity of the screening test.
B. A decrease in the calculated positive predictive value of the screening test.
C. An increase in the observed five-year survival for the target disease.
D. A decrease in the observed incidence rate of the target disease.
E. A decrease in the real mortality rate from the target disease.

Question 9
For many diseases, there is a critical point in time beyond which therapy is less effective. For most cancers, the critical point occurs when the primary tumor metastasizes. When must the critical point occur for screening to be effective?

A. Before the disease is detectable.
B. After the disease produces symptoms.
C. After the disease is detectable but before it produces symptoms.
D. After the disease is detectable and after it produces symptoms.
E. Before the disease is detectable and before it produces symptoms.

Question 10
Which one of the following statements related to length bias is true?

A. Rapidly progressing cases of disease are more likely to be detected during screening than are slowly progressing cases.
B. Slowly progressing cases of disease are more likely to be detected during screening than are rapidly progressing cases.
C. Slowly progressing cases of disease are less likely to be detected during screening than are rapidly progressing cases
D. Rapidly progressing cases tend to be detectable for a longer period of time than do slowly progressing cases.
E. Interval cases of disease tend to be more slowly progressing than are screen-detected cases.

SOLUTIONS

Solution to Question 1

Randomized controlled trials are considered the highest level of evidence because they evenly distribute the potentially confounding factors that can affect outcome [1-4]. Option D is the best response.

Randomization is necessary to reliably determine the effectiveness of medical interventions, including treatment, screening, and prevention, because these evaluations involve both an intervention group and a control group. However, evaluations of diagnostic accuracy do not involve an experimental group and a control group and therefore a randomized control trial is not appropriate (although randomization may be used to vary the order of imaging exams in a comparison study of accuracy). Prospective cohort studies are

considered the highest level of evidence for the evaluation of diagnostic accuracy and the second highest level, below randomized controlled trials, for the evaluation of effectiveness. Option A is not the best response. Case-control studies are considered one level below cohort studies because they are retrospective and more susceptible to bias. Case series have a lower level of evidence than cohort studies and case-control studies because the selection of cases is subject to greater sampling bias. Option C is not the best response. Expert opinion is considered the lowest level of evidence because it is prone to subjectivity and bias [1, 3]. Neither options B nor E is the best response. The relative levels of evidence are summarized in Table 1.

Table 1. Levels of evidence [5]

Level	Study design
1	Randomized controlled trial
2	Cohort study
3	Case control study
4	Case series
5	Expert opinion

Solution to Question 2

In a cohort study, subjects are selected on the basis of an exposure and followed prospectively for occurrence of a disease or other outcome of interest [2]. The observed relationship between the outcome and the disease is reported as a relative risk, which can be calculated as the ratio of the incidence of the disease among the exposed and the incidence of the disease among the unexposed. Using Table 2, the mathematical relationship is [a/(a+b)] / [c/(c+d)]. A well known cohort study related to radiology is the Katayama study of intravenous contrast media [5]. In this study, begun in 1986 while ionic contrast media was being replaced by nonionic contrast media throughout Japan, more than 300,000 patients received one of these forms of contrast media, as determined by the local radiology departments' policies, and were closely monitored for adverse drug reactions (ADRs). The incidence of ADRs was 3.1% in the group receiving nonionic contrast and 12.7% in the group receiving ionic contrast for a risk ratio of 0.25. Because the risk ratio was less than 1.0, nonionic contrast was considered protective against ADRs. A cohort study design is most useful for evaluating a new exposure that may be associated multiple outcomes, as in the preceding study of the new nonionic contrast media. If the subjects could have been randomly assigned to receive ionic or nonionic contrast media, an even stronger randomized controlled study might have been performed.

In a case-control study, cases and controls are selected on the basis of having and not having a disease or other outcome of interest, respectively [2]. The exposures of the cases and controls are then retrospectively assessed to determine their associations with the outcome. There is no way of knowing the incidence of disease among exposed and unexposed subjects in a case-control study because the number of subjects with and without disease is determined by the

investigator's selection criteria. Therefore, relative risk cannot be calculated. A different but similar number, the odds ratio, is used instead. Using Table 2, the odds ratio equals ad/bc. The stronger the association between the exposure and the disease, the higher the odds ratio; an odds ratio less than 1.0 implies a negative association. When the actual incidence of disease is low (much less than 1%), the odds ratio and the relative risk are essentially the same.

Blackmore used a case control study design to determine the relationship between cervical spine fractures and clinical predictors (exposures) of these fractures [6]. In this study, the clinical records of 472 patients, 168 with fractures and 304 without fractures, were retrospectively reviewed for the presence of 20 potential clinical predictors of these fractures. The strongest predictor was a focal neurological deficient, which had an odds ratio of 58. A case-control study design is most useful for evaluating a rare outcome that may be associated with many different types of exposures (or predictors), as in the preceding study of cervical spine fractures. A case-control study design is often the only practical design when there is a long latency between the exposure and the outcome and when the outcome is rare, as in the case of radiation-induced malignancy [7].

One major advantage of the cohort study design is that outcomes following exposure are assessed prospectively using well-defined and reproducible methods [2]. In a case-control study design, an outcome, such as death from a certain disease, is used to select the cases, and the exposure is assessed retrospectively through records that have already been assembled. Because of the temporal relationship between exposure and outcome, cohort studies may suggest more strongly a causal relationship between exposure and outcome than case control studies. Option C is the best response.

Because cohort studies are prospective and investigators must wait for the outcomes to occur, they take longer to complete than case-control studies, which are retrospective, and typically require searching records. Option A is not the best response. Both cohort and case-control studies are observational in nature and thus susceptible to bias. Option B is not the best response. In addition, both types of studies involve human subjects and therefore require IRB (institutional review board) approval [8]. Option D is not the best response. One disadvantage of the cohort study design is that only one type of exposure can be studied because subjects are selected on the basis of their exposure. Option E is not the best response.

Table 2. Exposure versus Disease (or other outcome).

	Disease Present	Disease Absent
Exposed	a	b
Not Exposed	c	d

Notes:
a = number of exposed subjects with disease
b = number of exposed subjects without disease
c = number of unexposed subjects with disease
d = number of unexposed subjects without disease

Solution to Question 3

The major advantage of the case-control study design over a cohort study design is that the former is more efficient for investigating rare outcomes than the latter because the cases are selected retrospectively on the basis of already having had experienced the rare outcome [2]. Option A is the best response.

However, a major disadvantage of the case-control study design is that the exposures are also determined retrospectively and thus may be recalled (or recorded) differently among the cases than among controls, thus creating recall bias [9]. Neither options B nor D is the best response. Both types of studies involve human subjects and thus require IRB approval [8]. Option C is not the best response. Both types of studies are observational in nature and susceptible to confounding. Option E is not the best response.

Solution to Question 4

Randomization is the most effective means of evenly distributing both know and unknown potentially confounding variables [2, 4]. Option E is the best response.

For observational studies, stratification, adjustment, and matching can all be used to help reduce the effect of potential confounders but cannot account for unknown confounders. None of options A, B, or C is the best response. Similarly, investigators cannot evenly distribute the confounders that are unknown among the study groups and they would likely introduce bias if they could influence group assignments. Option D is not the best response.

Solution to Question 5

In a case-control study, the relevant measure of association is the odds ratio [2]. In this example, the odds ratio equals 100x100/5x200, which equals 10. This means that smokers are about 10 times more likely to develop lung cancer than are non-smokers. Option D is the best response.

The prevalence of smoking in this case-control study is 74%. However, the prevalence of an exposure in the general population cannot be reliably determined from data case-control study because the cases and controls are selected on the basis of their outcomes and availability. Option A is not the best response. Similarly, the lung cancer rates among smokers and nonsmokers cannot be determined from this study. Outcome rates in case-control studies are generally much higher than are outcome rates in the general population because, by design, the cases are over-represented in the study population Neither option C, or E is

the best response. Finally, a case-control study can establish an association between exposure and outcome, but it cannot establish a causal link. Option B is not the best response.

Solution to Question 6

Because the purpose of screening is to reduce mortality from the target disease, the appropriate metric for the evaluation of its effectiveness is the ratio of the disease-specific mortality in the screened group to that in the control group. Comparisons of five-year survival overestimate screening effectiveness because of lead-time, length, and overdiagnosis biases [9-13].

Advancing the time of diagnosis will prolong the survival from the time of diagnosis even if it does not delay the time of death. The time interval between when the diagnosis would have occurred with screening and when it would have occurred without screening is known as the lead time. Option A is not the best response. Length bias pertains to the length of the detectable preclinical phase of the disease (DPCP). Because slowly progressing forms of disease have a longer DPCP than rapidly progressing forms, screening preferentially detects the former. Consequently, screen detected cases of disease would be expected to survive longer than clinically detected cases even after adjustment for lead-time. Option E is not the best response. Finally, screening may detect some preclinical cases of disease that would not have become clinically significant had they not been detected; such cases would inflate the 5-year survival. This phenomenon is known as overdiagnosis. Option D is the best response. A prospective study would not be affected by recall bias. Option B is not the best response. Selection bias occurs when subjects in the comparison are different for reasons other than the intervention under investigation. Randomization is the only reliable method of ensuring that two different groups are comparable. Furthermore, the screen detected cases in this comparison only occur in those complying with the screening program and compliant patients tend to have better outcomes. [14]. Option C is not the best option.

Comparisons of five-year survival are appropriate for determining treatment effectiveness in randomized clinical trials of treatment because in treatment trials, unlike screening trials, diagnosis occurs before the time of randomization [13]. Comparisons of five-year survival are inappropriate for determining screening effectiveness regardless of study design. Randomization by itself does not address the biases associated with early detection.

Solution to Question 7

The major purpose of randomization is to minimize the differences between subjects in the screened group verses those in the control group [2, 4, 9, 13]. Option D is the best response.

Ascertainment bias results from misclassification in the cause of death. If deaths from the disease of interest in the screened group are misclassified as death from another cause, then the risk ratio is biased in favor of screening. However, if deaths from the disease of interest in the control group are misclassified as death

from another cause, then the risk ratio is biased against screening [9, 13]. Option A is not the best response. Once subjects are randomized they should be analyzed according to their assignment, regardless of whether they comply with their assignment. This is known as the "intent to treat" principle [4]. Option E is not the best response.

Lack of compliance with the randomized assignment in either the screened group or control group results in an underestimation of the screening effect. For example, if subjects who are randomly assigned to the screening group are noncompliant and do not obtained the screening, the screening effect will be underestimated. Conversely, if subjects in the control group decide to get screening, it will also tend to reduce the differences between the screening and control groups, resulting in an underestimation. Neither option B nor C is the best response.

Solution to Question 8

Overdiagnosis is the diagnosis of a condition that would not have become clinically significant if it had not been detected by screening [9, 15]. Two types of overdiagnosis have been recognized. Type I refers to conditions that do not progress. For example, many cases of carcinoma-in-situ do not progress to invasive cancer (the "gold standard" of pathology is not truly golden). Type II refers to conditions that would have progressed if the individual had not died from another cause. For example, elderly men with microscopic prostate cancer usually die from other causes long before their cancers would have become clinically significant. Either type of overdiagnosis increases the observed five-year survival because it adds cases without real disease to the numerator and denominator of the survival statistic [9, 13, 15]. Option C is the best response.

Overdiagnosis leads to an increase in the calculated sensitivity and positive predictive value of the screening test because it adds cases without real disease to the numerator and denominator of these accuracy statistics. Neither option A nor B is the best response. Overdiagnosis leads to an increase in the observed incidence rate of the target disease because overdiagnosed cases are falsely counted as real cases of disease. Option D is not the best response. Overdiagnosis does not affect the real mortality rate from the target disease and this is one major justification for using disease-specific mortality as the major endpoint in randomized controlled trials of screening. Option E is not the best response.

Solution to Question 9

For many diseases, there is a critical point in time beyond which therapy is less effective [9, 13] For most cancers, the critical point occurs when the primary tumor metastasizes. For screening to be effective, the critical point must occur within the detectable preclinical phase disease (DPCP) [9, 13]. Option C is the best response. If the critical point occurs before the detectable preclinical phase of disease, then screening is futile. Options A and E are not the best responses. If the

critical point occurs after the detectable preclinical phase of disease, then screening is unnecessary. Options B and D are not the best responses.

.

Solution to Question 10

Slowly progressing cases of disease are more likely to be detected during screening because they exist in the detectable preclinical phase (DPCP) for a longer period of time than do rapidly progressing cases [9, 11, 13]. Option B is the best response.

The rapidity of disease progression is inversely related to the length of the DPCP. Rapidly progressing cases of disease are less likely to be detected during screening than are slowly progressing cases because the former exist in the DPCP for a shorter period of time. None of options A, C, or D, is the best response. Interval cases, which surface clinically after a negative screening exam, tend to be more rapidly progressing than are screen-detected cases. Option E is not the best response.

REFERENCES

1. Blackmore, C.C., *The challenge of clinical radiology research.* AJR 2001; **176**(2):327-331.
2. Blackmore, C.C. and P. Cummings, *Observational studies in radiology.* AJR 2004; **183**(5):1203-1208.
3. Centre for Evidence-Based Medicine. *Levels of evidence and grades of recommendation.* [Available at: http://www.cebm.net/levels_of_evidence.asp. Accessed May 2007].
4. Stolberg, H.O., G. Norman, and I. Trop, *Randomized controlled trials.* AJR 2004; **183**(6):1539-1544.
5. Katayama, H., et al., *Adverse reactions to ionic and nonionic contrast media. A report from the Japanese Committee on the Safety of Contrast Media.* Radiology 1990; **175**(3):621-628.
6. Blackmore, C.C., et al., *Cervical spine imaging in patients with trauma: determination of fracture risk to optimize use.* Radiology 1999; **211**(3):759-765.
7. Macmahon, B., *Prenatal x-ray exposure and childhood cancer.* J Natl Cancer Inst 1962; **28**:1173-1191.
8. Cooper, J.A., *Responsible conduct of radiology research: part II. Regulatory requirements for human research.* Radiology 2005; **236**(3):748-752.
9. Black, W.C. and H.G. Welch, *Screening for disease.* AJR 1997; **168**(1):3-11.
10. Blackmore, C.C. and L.S. Medina, *Evidence-based radiology and the ACR Appropriateness Criteria.* J Am Coll Radiol 2006; **3**(7):505-509.
11. Herman, C.R., et al., *Screening for preclinical disease: test and disease characteristics.* AJR 2002; **179**(4):825-831.

12. Welch, H.G., L.M. Schwartz, and S. Woloshin, *Are increasing 5-year survival rates evidence of success against cancer?* JAMA 2000; **283**(22):2975-2978.
13. Black, W.C., *Randomized Clinical Trials for Cancer Screening: Rationale and Design Considerations for Imaging Tests.* J Clin Oncol 2006. **24**(20):3252-3260.
14. Fletcher RD, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials. Third edition. Philadelphia: Lippincott Williams & Wilkins, 1996:180.
15. Black, W.C., *Overdiagnosis: An underrecognized cause of confusion and harm in cancer screening.* J Natl Cancer Inst 2000; **92**(16):1280-1282.