Opinion

M. G. Myriam Hunink, MD, PhD

Gabriel P. Krestin, MD, PhD

See also the editorial by Jarvik in this issue.

Study Design for Concurrent Development, Assessment, and Implementation of New Diagnostic Imaging Technology[1]

With current constraints on health care resources and emphasis on value for money, new diagnostic imaging technologies must be assessed and their value demonstrated. The state of the art in the field of diagnostic imaging technology assessment advocates a hierarchical step-by-step approach. Although rigorous, such a hierarchical assessment is time-consuming, and, given the current rapid advances in technology, results are often too late to influence management and policy decisions. The purpose of this article is to discuss a study design in which development, assessment, and implementation of new diagnostic imaging technology take place concurrently in one integrated process. An empirically based pragmatic study design is proposed for imaging technology assessment. To minimize bias and enable comparison with current technology, a randomized controlled design is used whenever feasible and ethical. Outcome measures should reflect the clinical decision-making process based on imaging information and acceptance of the new test. Outcome measures can include additional imaging studies requested, costs of diagnostic work-up and treatment, physicians' confidence in therapeutic decision making, recruitment rate, and patient outcome measures related to the clinical problem. The key feature of the proposed study design is analysis of trends in outcome measures over time.

The rapid advances in diagnostic imaging technology make a thorough assessment of new technology prior to its implementation a challenge. With the current constraints on health care resources and the emphasis on value for money, we feel compelled to assess new technologies and demonstrate their value prior to their widespread use in day-to-day practice. A hierarchical approach to the assessment of new diagnostic imaging technology has been advocated and used by technology assessors in the field (Fig 1) (1–12). This approach entails assessments of technical performance; diagnostic performance; diagnostic, therapeutic, and prognostic effect; effectiveness (patient and societal outcomes); and cost-effectiveness of the new technology.

The hierarchical approach is rigorous and scientifically sound, but when performed sequentially it is also time-consuming. We often find that by the time we have performed the advocated studies, the results of such an assessment are no longer useful. With the rapid advances in diagnostic imaging technology, either the new test has been implemented in practice because of the "beautiful images" it produces, or it has been discarded in lieu of an even newer technique that clearly surpasses the one we have been evaluating. In fact, new diagnostic imaging technologies are often implemented as soon as they are available, on the basis of subjective experience with a limited number of cases. An additional problem is that for each modality, new technologic developments, such as new magnetic resonance (MR) sequences or updated postprocessing software, are constantly implemented on the basis of subjective expectations of their usefulness.

The goal of developing and assessing new diagnostic imaging technology is presumably
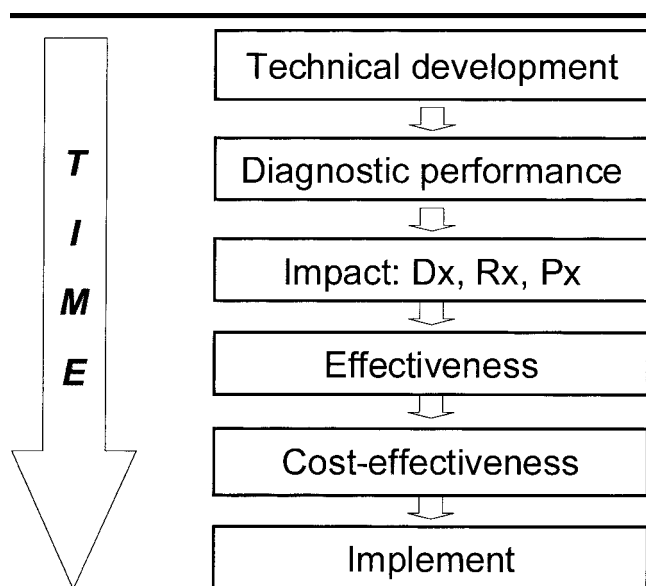
Figure 1. Schematic shows the hierarchical approach to development, assessment, and implementation of new diagnostic imaging technology. *Dx* = diagnosis, *Px* = prognosis, *Rx* = therapy.

to implement valuable and affordable new technology in a timely fashion to attain our fundamental goal of improving health. To achieve this goal we feel the need to perform thorough scientifically based assessment studies of new diagnostic imaging technology prior to its implementation, to demonstrate that the technology provides value for money (13). At the same time, we feel the need to rapidly implement new diagnostic imaging technology, which is why we tend to base our judgment of the value of new technology on subjective experience with a limited number of cases. As demonstrated in Figure 2, these considerations give rise to two apparently disparate pathways to our goal (14,15). The diagram represents a closed circle, which symbolizes the conflict that exists between the two pathways. As long as the circle remains closed—that is, as long as the conflict remains unresolved—tension will persist. So, how do we resolve this conflict?

In this article, we address the conflict between the perceived need for a hierarchical assessment of new diagnostic imaging technology prior to its implementation versus the perceived need to rapidly implement new technology after a subjective evaluation based on our experience with a limited number of cases. To resolve the conflict, we propose a study design that integrates the development, assessment, and implementation of new diagnostic imaging technology.

CHALLENGING THE UNDERLYING ASSUMPTIONS

To address the diagramed conflict, we must identify the underlying assumptions and find evidence that could be used to either support or weaken the assumptions. Let us analyze each arrow in Figure 2.

(A) Only Thorough Hierarchical Assessment Will Provide the Necessary and Relevant Information
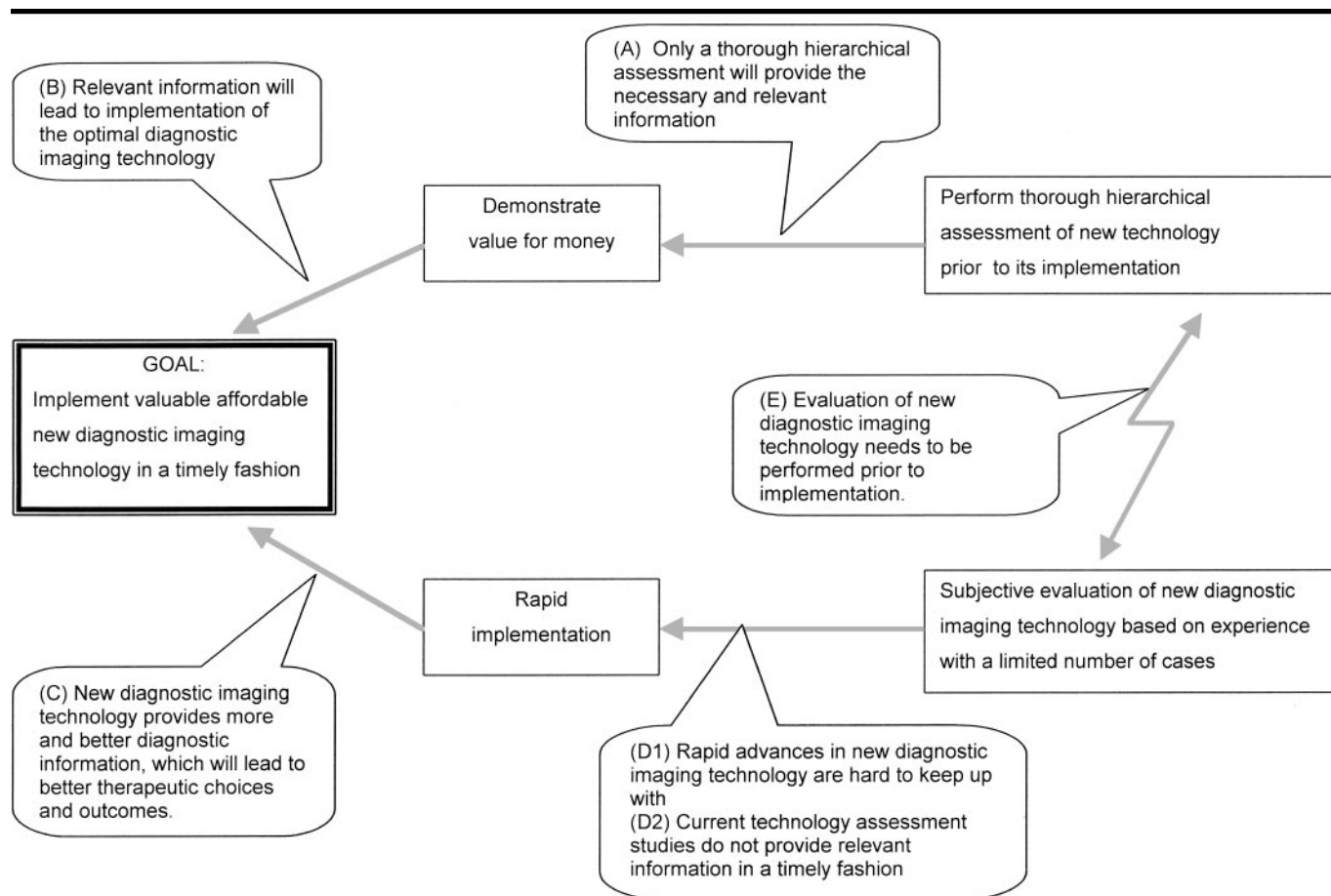
Traditionally, the evaluation of new diagnostic imaging technologies has focused on determining pairs of sensitivity and specificity (or a receiver operating characteristic curve) values in comparison to a reference standard. This entails performance of both the new test and the reference standard test in all patients in a cohort study and determination of the probability of abnormal and normal findings conditional on disease or no disease (sensitivity and specificity, respectively) or, alternatively, the predicted probability of disease conditional on the test result. The reference standard, however, is not flawless in helping distinguish individuals with disease from those without. For example, intraarterial x-ray angiography (an invasive procedure involving arterial catheterization) is generally considered to be the reference standard for vascular disease but has been

shown to result in missed patent runoff vessels (16,17). Because computed tomographic (CT) angiography and MR angiography (18–20) provide three-dimensional information, they may surpass intraarterial angiography. A comparison of CT angiography or MR angiography with intraarterial angiography as the reference standard will, therefore, lead to underestimation of the sensitivity and specificity of these new tests because intraarterial angiography is defined as the reference standard, or perfect, test. Furthermore, for some innovative technologies such as molecular imaging, an appropriate reference standard simply does not exist because the index test is an attempt to diagnose a condition that is unidentifiable with existing technology (eg, cancer in a very early phase).

In addition, if the reference standard examination is invasive we cannot, on ethical grounds, perform it in all cases. This implies that the new test will be verified in selected cases only, and this, too, may lead to biased estimates of sensitivity and specificity (21–23). For example, in potential living kidney donors, surgical findings may be considered the best reference standard, but surgical findings will be available only for those kidneys removed for transplantation. Commonly, we revert to using a combination of findings as the reference standard (24,25). Such an approach inevitably leads to biases related to the mixed reference standard and a reference that is not independent of the tests being evaluated (21–23). Alternatively, we can use a mathematical correction method that adjusts for the potential bias resulting from selected verification (26–30). This method does, however, require the assumption that the predictive value of a test is unaffected by verification bias.

Furthermore, although sensitivity and specificity may be useful performance parameters in the initial evaluation of a new test, they seldom provide the information we need to decide whether the new diagnostic strategy should be implemented. A result indicating that sensitivity and specificity are both, say, 95% is difficult to translate into a meaningful clinical decision. To decide whether the new diagnostic strategy should actually replace the current strategy requires studies to evaluate the effect on decision making, patient outcomes, and costs.

Decision analysis is a useful tool for estimating the effect on cost and effectiveness outcomes and for evaluating diverse strategies (31–33). Advantages of decision analysis are the ability to inte-

Figure 2. Diagram illustrates the conflict between performing a hierarchical assessment versus subjective evaluation of new diagnostic imaging technology.

grate all the available evidence and values, model a wide range of strategies, and explore the effect of uncertainty on the decision. Decision analysis does, however, have limitations. It is time-consuming to perform properly, it relies on data from multiple heterogenous sources, it requires assumptions to make the problem tractable, and it has limited impact on everyday clinical practice. A large randomized controlled trial (RCT) with long-term follow-up of quality-of-life and survival measures would be scientifically rigorous, but, because the differences in quality of life and survival across diagnostic strategies are generally small (34), such studies must be large, take a long time to perform, and require substantial resources. One can question whether the additional information obtained from such trials justifies the research resources used (35).

(B) Relevant Information Will Lead to Implementation of Optimal Diagnostic Imaging Technology

Even if we have the resources to perform a thorough evaluation of a new di-

agnostic imaging technology, implementation of the study results can pose a problem. After performing a cohort study in which both the old test and the new test are performed in all patients, we subsequently often find ourselves performing both tests because that is what clinicians have become familiar with. Both physicians and patients assume that more imaging studies will lead to better diagnostic information. Furthermore, more tests yielding the same findings increase confidence in the diagnosis. Thus, physicians will tend to want to use all the possible diagnostic imaging technologies available. Four of five times, implementation of new technology implies that instead of the new technique replacing the old technique, the new technique is performed in addition to the old one (36).

(C) New Imaging Technology Provides More and Better Diagnostic Information That Will Lead to Better Therapeutic Choices and Outcomes

The tremendous increase in the use of imaging procedures suggests that physi-

cians and patients believe that more and better diagnostic imaging technology leads to more and better diagnostic information, which in turn leads to optimal therapeutic choices and outcomes. The connection between more diagnostic imaging information and patient outcomes, however, is difficult to demonstrate because of the multiple intervening paths and steps (36). Although the choice of treatment is commonly influenced by imaging findings, a change from a currently used diagnostic strategy to one that uses a newer technology will generally yield only small benefits to the patient (34). New imaging technology can, however, reduce the convalescence period through the use of procedures, both diagnostic and therapeutic, that are less invasive.

(D1) Rapid Advances in New Diagnostic Imaging Technology Are Hard to Keep Up With

Rapid technologic advances are a fact. Not only do we accept this fact, but we are also generally content with the tech-

nologic advances, since they bring with them the possibility of better health care and a higher standard of living. New technology is becoming more widely available, there is a continual increase in the expertise of those using new technology, and there is continual development of the technology. The pace of the advances are mind-boggling, and we must do everything we can to cope with the technologic changes. In effect, this requires efficient methods of research, which implies a focus on providing the information relevant to the clinical decision-making process.

(D2) Current Technology Assessment Studies Do Not Provide Relevant Information in a Timely Fashion

In this respect we should ask ourselves: What exactly is the relevant information in this context? To decide whether a new technology is valuable and affordable, do we really need to go through the whole gamut of the test's reproducibility; sensitivity; specificity; receiver operating characteristic analysis; effect on diagnostic, therapeutic, and prognostic thinking; patient benefit from use of the test; and cost-effectiveness from a societal perspective? This is probably not so. In fact, the only evidence we really need is that use of the new diagnostic strategy facilitates the clinical decision-making process without compromising patient outcomes and that we can afford to pay for it.

Sometimes a new technology is so obviously better, simpler, less risky, and less expensive than the established technology that an extensive assessment is unnecessary: A balance sheet of the pros and cons, taking into account the various dimensions of the decision, can suffice (31). For example, head CT did not need to be extensively assessed in comparison to pneumoencephalography to demonstrate its superiority. At the same time, statements of obvious superiority must be interpreted with caution: The advocate of a new technology is usually someone with his or her own biases and agenda. Disillusionment frequently follows an initially optimistic introduction of new technology. Currently, new technology is generally only marginally better than what we have available already, and the associated price tag may not be justified. Commonly, the new diagnostic strategy has both advantages and disadvantages in comparison with the established tests, in which case an assessment of its effect on the clinical process is warranted.

(E) Evaluation of New Diagnostic Imaging Technology Must Be Performed before Implementation

With the hierarchical approach, evaluation of a new technology must precede its implementation. Although this appears to make good sense, one can question its necessity and desirability. In fact, there may be a study design that allows simultaneous evaluation of the new technology and implementation of the optimal diagnostic strategy. In other words, performance of the study ideally not only helps assess the new technology but at the same time improves clinical practice through some form of a self-organizing feedback system (37). In a self-organizing feedback system, adjustments of physician ordering behavior, of the technology used, and of the interpretation take place as a result of feedback loops.

The above list of problems and questionable assumptions suggests that we should think of a study design for the evaluation of new diagnostic imaging technology that gets away from our focus on cohort validation studies to evaluate sensitivity and specificity. So, is there a study design that can address the issues listed?

CRITERIA FOR THE ASSESSMENT AND IMPLEMENTATION OF NEW DIAGNOSTIC IMAGING TECHNOLOGY

The goal of a study to evaluate new diagnostic imaging technology is to help us decide whether the new technology provides value for money and to enable implementation of the new technology in a timely fashion. To achieve this goal the following criteria must be met: *(a)* The study should demonstrate that the new diagnostic strategy facilitates the clinical decision-making process. *(b)* The process of data collection should be integrated into clinical practice. *(c)* By the end of the study, the optimal diagnostic strategy should have been implemented in clinical practice through a self-organizing feedback system. *(d)* The study should be fast and require only limited resources. *(e)* The study should demonstrate that the new diagnostic strategy leads to equivalent or better outcomes for the patient. *(f)* The study should demonstrate that the cost of the new diagnostic strategy is either the same as or less than that of the strategy it is replacing or that the benefits to the patient or cost savings

downstream are so large that we can afford to pay for the new test.

STUDY DESIGN: AN EMPIRICAL RCT

We propose assessing diagnostic imaging technology with an empirically based study design. To minimize the bias introduced by extraneous factors and ensure an unbiased comparison with a control group, we propose using an RCT design whenever feasible and ethical. Outcomes measured would need to reflect the global goal of performing the imaging test in the clinical context under consideration. An essential feature of our proposed study design is that we propose taking into account the dimension "time" when analyzing the outcome measures. We will discuss each of these issues in turn.

Empirical Study

Our goal of timely implementation of valuable new diagnostic imaging technology could potentially be achieved if we strive for concurrent implementation and assessment of the new technology. That is, the process of performing the trial should lead not only to evaluation of the new diagnostic strategy in comparison with the old strategy but also to implementation of the new strategy if it is better. The trial design should, in our opinion, be empirically based and pragmatic; that is, the trial should be integrated into clinical practice rather than be implemented in a strictly controlled, but probably unrealistic, experimental setting.

In general, current clinical practice would be used as the control strategy for the study. We recognize that current clinical practice is, in many instances, unstructured and that documentation of relevant information may be lacking. Clearly, sloppy clinical practice will not suffice as the control strategy. Ideally, performance of the study will help structure and streamline clinical practice, initiate collection of relevant information, and enhance communication between physicians mutually and between physicians and their patients.

Randomization across Diagnostic Strategies

As with all studies, the inclusion and exclusion criteria should guarantee selection of a representative group of patients for the diagnostic problem under consid-

eration. The tradeoff is between recruiting a fairly homogeneous group of patients that is clearly identifiable versus ensuring that the study is sufficiently generalizable (38).

After recruiting patients that fulfill the inclusion criteria, we propose random assignment of patients to the new diagnostic strategy or to the old strategy. For example, if we want to evaluate CT angiography versus intraarterial angiography for peripheral arterial disease, we would randomly assign patients with peripheral arterial disease to undergo either initial CT angiography or intraarterial digital subtraction angiography (DSA) (Fig 3). In this empirical design, we would allow for further evaluation with other imaging technology should the need arise. For example, if CT angiography does not provide sufficient diagnostic information to establish a therapeutic plan, the patient would undergo (selective) intraarterial angiography.

Randomizing Discordant Test Results

Randomization after inclusion and prior to the initial diagnostic test is the most straightforward practical approach. It may not, however, always be the most efficient approach, because a large proportion of the patients will not contribute to a difference in outcome. An alternative design is to study all patients with both tests (both CT angiography and intraarterial DSA) and randomly assign only those patients with discordant test results (ie, CT and intraarterial DSA yield discrepant results). Although this may be a more efficient design (ie, fewer patients required for the study), the logistics of such a design are far more complicated, and the results may lead to interpretation problems.

For example, in this clinical setting the test result is not a dichotomous yes/no result but rather a treatment choice out of a range of possible treatments, including bypass surgery with various possible proximal and distal anastomotic sites, percutaneous transluminal angioplasty or stent placement at various sites, combinations of surgical and percutaneous procedures, or no revascularization procedure. The treatment choice is further influenced by local preferences and expertise and by the patient's preferences. Thus, what constitutes a discrepancy in the test results is not easily defined. Furthermore, if both tests have been performed, information from the test that is not supposed to influence management
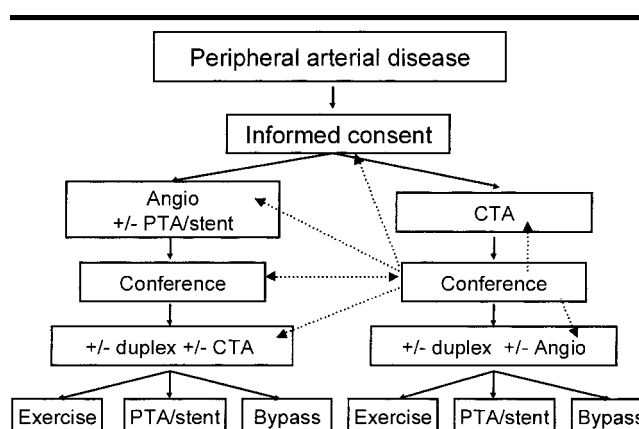
Figure 3. Schematic shows study design for an empirical RCT to compare CT angiography *(CTA)* with intraarterial DSA *(Angio)* for pretreatment diagnostic imaging work-up of peripheral arterial disease. After randomization, either CT angiography or intraarterial angiography (in a few cases followed immediately by percutaneous transluminal angioplasty or stent placement *[+/-PTA/stent]*) is performed. The imaging findings are discussed at the vascular conference. The imaging work-up is required to help determine the optimal treatment: percutaneous transluminal angioplasty, stent placement, bypass surgery, or exercise therapy. The dotted arrows indicate where subjective experience is expected to influence the clinical process. *duplex* = duplex ultrasonography, +/- = with or without.

will invariably become known to the treating physicians in a proportion of cases, in part indirectly because the treating physician will know that the other test yielded a discrepant result (otherwise this patient would not have been randomly assigned) and in part because information that is withheld for research reasons can always be requested on ethical grounds. Finally, interpretation of the results may be hampered by the fact that the diagnostic work-up itself may influence not only the choice of treatment but also the manner in which the treatment is performed (eg, if a CT angiographic image is available to guide catheterization, the procedure may be performed differently than if no prior noninvasive imaging information is available).

Randomization between Providing Test Results versus Not Providing Test Results

An alternative to randomizing across diagnostic strategies would be to use the new diagnostic strategy in all patients but only provide the test results to the treating physician and patient for those randomly assigned to the new diagnostic strategy. This approach facilitates recruitment of patients because all patients will be offered the new technology, which tends to increase their willingness to participate. Furthermore, this approach could potentially reduce a biased follow-up if pa-

tients are more willing to respond to questionnaires after participation in a new diagnostic strategy. The authors of one published study (39) used this approach to determine the value of MR imaging for help in assessing fetal-pelvic proportions in breech presentation and demonstrated that the information obtained with MR imaging increased the number of successful vaginal deliveries and reduced the number of emergency cesarean sections.

The main caveat of this approach is similar to that of randomization of discrepant test results. Information from the new test will invariably become known to the treating physicians in a proportion of cases in the group where it was not supposed to influence management according to the study protocol. Part of the problem is the logistics of keeping test results secret, but this can be overcome with the correct procedures in place. More important, information withheld for research reasons can be requested for an individual patient on ethical grounds, and if the diagnostic test has been performed already it is a small step for the treating physician to insist on getting the information. Finally, if the new test is associated with a risk, even a small one, it may be considered unethical to perform such a test when the test result does not have any influence on the care of the patient.
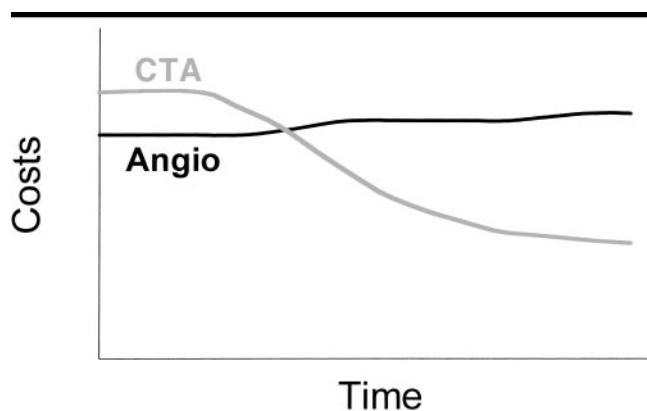
Figure 4. Graph shows expected trends in costs of diagnostic work-up and treatment. Patients undergoing the experimental strategy, in this example CT angiography *(CTA)*, may still need additional work-up with the control strategy, intraarterial DSA *(Angio)*. As physicians gain more experience with the new technology over time, additional examinations will probably be required less often, and the costs of the work-up with the experimental strategy will decrease. Patients undergoing the control strategy may require additional work-up, which may, over time, be performed more and more with the experimental strategy.

A Nonrandomized Design

Although in general we would recommend an RCT design, some situations may justify a modified design. Sometimes the potential study population is so small that randomization does not seem feasible within a limited time frame. Sometimes the new diagnostic strategy is far safer or far less burdensome to the patient than the currently used diagnostic test. Sometimes preliminary reports from other institutions have already demonstrated good results with the new diagnostic strategy, making randomization unethical or impossible because of patient preferences and physician bias.

A nonrandomized design would imply that, first, the new test would be performed in addition to the current work-up to "fine tune" the technical details of performing the examination and to familiarize physicians (both radiologists and clinicians) with the new diagnostic imaging technology. During this phase the current work-up should be established in protocol form and documented. Approximately 10–20 cases would probably suffice for this first phase. Next, the new diagnostic strategy should be implemented as the initial test. Physicians are given the option of requesting the previously used test if they believe it is necessary to complete the work-up, but such requests are recorded and tracked. The process is documented over time as physicians become familiar with the new test.

Outcome Measures and Trends over Time

As mentioned, we consider the trial to be a vehicle to implementation of the best diagnostic imaging technology. We propose evaluating the implementation process by including outcome measures that reflect the clinical decision-making process based on the imaging information and acceptance of the new test. Because physicians' experience with and opinions of the test change over time, all outcome measures would have to be tracked over time. Crucial to our proposed study design is to measure outcomes that reflect the learning curve and acceptance of the new diagnostic strategy over time. Thus, taking into account the dimension "time" is an essential feature of our proposed study design.

Furthermore, the outcome measures should focus on the goal of performing the diagnostic trial and should therefore measure the clinical decision-making process. Outcome measures that reflect the goal of the trial include the probability that additional studies would be ordered, the costs of the diagnostic work-up and treatment, the physicians' confidence in therapeutic decision making, patient outcome measures related to the clinical problem at presentation, and, in the event of a randomized trial, the recruitment rate.

Percentage (and type) of additional requested tests and trends over time.—One would expect that this percentage is practically constant over time for the control

(ie, currently used) diagnostic work-up strategy and variable over time for the experimental strategy. For example, in the course of an implementation study of CT angiography compared with intraarterial DSA for peripheral arterial disease, one would expect the percentage of requested intraarterial DSA procedures following CT angiography to be high initially (probably about 90%) and to decrease over time as interventional radiologists and vascular surgeons become more familiar with this technology for this particular indication. In other words, the proportion of intraarterial DSA procedures requested would reflect the learning curve of performing and interpreting CT angiographic images. Furthermore, one would expect that after initial CT angiography, a requested angiogram could be limited to a selective study. It is even conceivable that with the introduction of CT angiography, physicians will request CT angiography after intraarterial DSA to solve specific diagnostic problems. The latter would demonstrate a shift in the control strategy over time that would support the use of the new technology.

Costs of diagnostic work-up and treatment and trends over time.—Resource utilization and the associated costs of the diagnostic work-up and treatment should be recorded, and the trends over time should be calculated (Fig 4). As indicated, it is likely that crossovers will occur, in that patients undergoing the experimental strategy may still need additional work-up with (part of) the control strategy. In addition, patients undergoing the control strategy may require additional work-up, which may, over time, be performed more and more with the experimental strategy. If the diagnostic strategy influences the treatment approach, this effect would need to be estimated. For example, the availability of CT angiographic images prior to a combined intraarterial DSA–angioplasty procedure could potentially lead to a more focused selective diagnostic procedure and a more direct approach to the interventional procedure.

Medical costs can be estimated by multiplying resource utilization and the cost per unit of resource. The cost per unit of resource—that is, per diagnostic or therapeutic procedure—should be determined with cost accounting that takes into consideration the initial investment of equipment, investments during use, maintenance, years of use, discounting, number of procedures per year, personnel costs, room rent, housekeeping, administration, and other overhead costs (13,31).
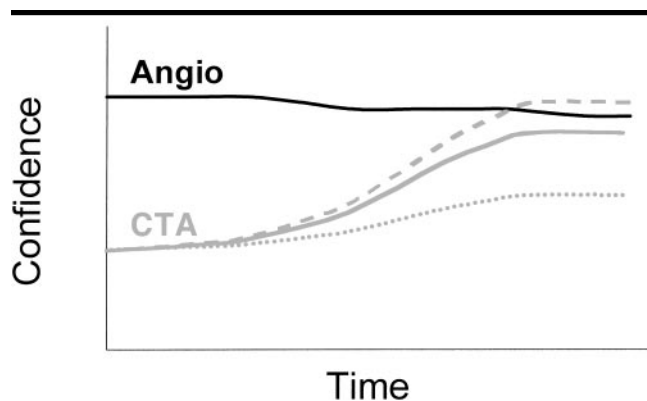
Figure 5. Graph shows expected trends in confidence in decision making. The physicians' confidence may be measured by using a rating scale or visual analogue scale. One would expect that their confidence is constant over time for the control strategy (intraarterial DSA *[Angio]* in this example). For the experimental strategy of CT angiography *(CTA),* one would expect their confidence initially to be lower than that for the control strategy and to increase over time as they become more familiar with the new technology.

Figure 6. Graph shows the recruitment rate as a function of time since the start of the study. As long as the two strategies are considered equivalent, physicians will be comfortable with patients being randomly assigned to treatment. If, over time, it becomes clear from clinical experience that use of one of the diagnostic imaging strategies is preferable, physicians will be reluctant to recruit patients for the trial. Despite strict inclusion and exclusion criteria, a subtle change in those considered eligible for the trial may take place. A criterion for stopping the study (arrow) could be based on the percentage of potentially eligible patients recruited for the study.
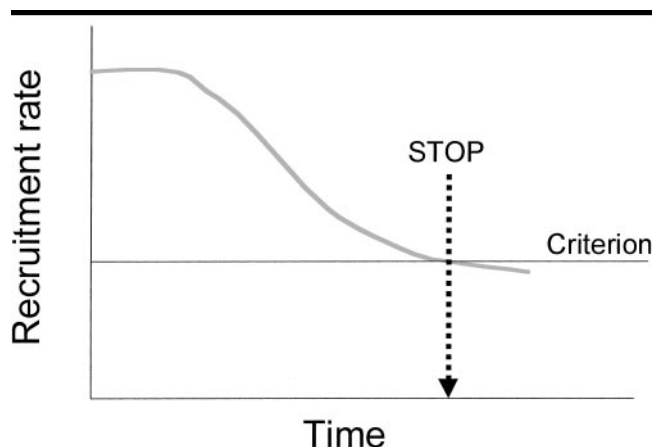
Confidence in therapeutic decision making and trends over time.—The diagnostic and therapeutic decisions made and the physicians' confidence with their therapeutic decision making are additional useful outcome measures (Fig 5). The physicians' confidence may be measured by using a rating scale or a visual analogue scale (40,41). One would expect that their confidence is constant over time for the control (currently used) diagnostic work-up strategy. For the experimental strategy one would expect their confidence initially to be lower than that of the control strategy and to increase over time as they become more familiar with the new technology. Confidence could, however, decrease over time if initial expectations of the new technique are not fulfilled as the limitations of the technology become apparent.

Patient outcomes.—To track whether potentially treatable disease is inadvertently missed when the experimental strategy is used, patients must be followed up for a period of time during which the false-negative outcomes would be apparent. For peripheral arterial disease, for example, 6 months would be appropriate. Patients with persistent symptoms would return within this period. During follow-up, patient outcomes should be assessed by researchers blinded to the allocated strategy. Patient outcome measures should include survival, symptoms, quality-of-life measures, and events, including the percentage of patients that cross over from initial conservative therapy to revascularization because of persistent symptoms (42,43). If CT angiography does not fail to demonstrate treatable disease, we would expect patient

outcomes and the percentage of crossovers to be the same for the two strategies.

The recruitment rate as function of time since the start of the study.—One of the outcome measures in our proposed study design is the percentage of eligible patients actually recruited for the study as a function of time since the start of the study (Fig 6). As long as the two strategies are considered to be equivalent, physicians will be comfortable with random assignment of patients across the diagnostic work-up strategies. Any trend identified in the proportion of patients recruited should be interpreted in the context of all factors that may influence recruitment. For example, during the course of the trial physicians may become reluctant to recruit patients without palpable femoral pulses and may instead request CT angiography. Despite strict inclusion and exclusion criteria, a subtle change in those considered eligible for the trial may occur. Furthermore, patients may pick up subtle nonverbal information during the informed consent procedure and, as a result, be unwilling to participate.

Both physicians' reluctance to recruit patients and patients' refusal to participate will be noticable in the recruitment rate. Although a change in recruitment rate will not change the internal validity of the trial, the generalizability of the results must be considered in light of the changing indication for the procedures evaluated. In accordance with the Con-

solidated Standards of Reporting Trials, or CONSORT, statement for the reporting of RCTs (44), we propose tracking the number of potentially eligible, the number actually eligible, and the number recruited patients over time. A criterion for stopping the study could be based on the percentage of potentially eligible patients recruited for the study. For example, if at the outset 70% of potentially eligible patients are recruited, and after 6 months only 20% are recruited, then the subjective experience of the physicians will probably have influenced recruitment, provided, of course, that other logistic problems have not arisen.

Data Collection

The data that would need to be collected in the context of the described trial for peripheral arterial disease include the following:

1. Date of presentation, date of completion of the diagnostic work-up (ie, date of last imaging study), diagnostic tests performed, date and type of the treatment decision, date of treatment, date that normal activities are resumed, date of recurrence of symptoms and severity of those symptoms, and type of events.

2. All patients with peripheral arterial disease evaluated by the vascular surgery staff would need to be registered and baseline characteristics recorded to enable the evaluation of generalizability
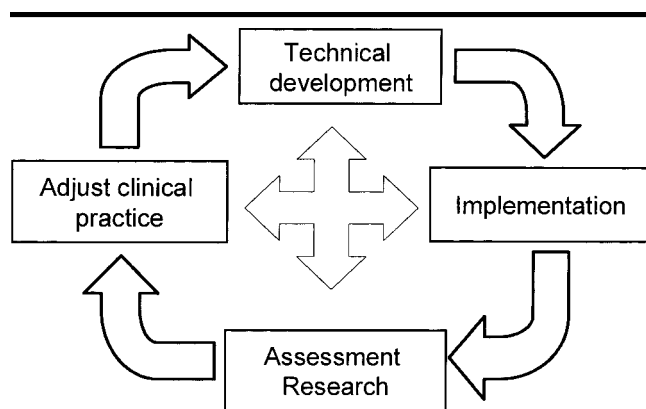
Figure 7. Diagram illustrates the concept of an interwoven, circular, self-organizing approach to development, assessment, and implementation of new diagnostic imaging technology.

and the proportion eligible and proportion randomly assigned.

3. Patient characteristics and diagnostic test results should be documented, including age, sex, symptoms (according to the Rutherford classification), and findings on CT and/or intraarterial DSA images (including lesion severity, lesion length, lesion location, and runoff). These variables are used in evaluation of the randomization process and to help analyze temporal changes in clinical decisions made.

4. If reproducibility of interpretation is an issue, the images should be interpreted by at least two independent observers.

5. Diagnostic and therapeutic decisions made must be tracked for each patient during the vascular conference, as must confidence in the decision. The latter can be recorded on a rating scale from zero to 10, with zero indicating no confidence and 10 indicating full confidence in the decision (41).

6. Quality of life should be measured at baseline, 6 weeks, 3 months, and 6 months, with a general descriptive instrument (eg, MOS 36-Item Short-Form Health Survey), a disease-specific descriptive instrument (eg, the VascuQol), and an evaluative instrument providing general population values (eg, the EuroQol) (45–48).

7. Resource utilization of diagnostic procedures must be tracked and a detailed cost analysis should be performed for all procedures that are important (eg, CT and DSA) by recording the time required to perform the procedure, the personnel involved, and the materials needed. Other costs per unit of resource can be analyzed independently of patient data collection.

8. Resource utilization for therapeutic procedures should be tracked with case record forms and complemented with data from the hospital information system and the patient questionnaires.

Data Analysis

Analogous to the intention-to-treat principle for analysis of therapeutic RCTs, the data of a diagnostic RCT should be analyzed on an intention-to-diagnose-and-treat basis. This implies that once a patient has been randomly assigned, he or she will remain in the assigned group for the analysis irrespective of whether crossover occurred to the other strategy and of whether follow-up was complete or not. The underlying concept is that patients that cross over and patients lost to follow-up are part and parcel of the strategy.

Essential in the analysis of data from a study that involves concurrent implementation and evaluation of a new diagnostic strategy is determination of trends among outcome measures over time. By including time since the start of the study in the analysis, we have a surrogate for the learning curve of physicians and we can adjust for this effect. The analysis can be performed pragmatically by using multivariable regression analysis to model costs of the diagnostic work-up, costs of the diagnostic work-up plus treatment, and physicians' confidence in the therapeutic decision. Multivariable logistic regression analysis can be used to model the probability of additional examinations requested and the recruitment rate. In both models, the independent variables include the diagnostic strategy; time since the start of the study; an interaction term between these two; and covariates such as symptoms, age, and sex. In a randomized design, the focus of the analysis would be to demonstrate a difference between the two groups, taking into account the trend over time. In the absence of a concurrent randomized control group, the analysis of trends over time is the primary focus of the analysis.

Randomization and selected use of the traditional test does not preclude the possibility of obtaining estimates of the sensitivity and specificity of a new test in comparison with those of the traditional test as reference standard. Because the test results are verified only in selected cases, we would need to use a method that adjusts for the associated possible verification bias (26–29). This requires development of a logistic regression model to predict the probability of verification as a function of age, sex, disease severity, and time since start of the study. The probability of verification is then used to adjust the data and calculate corrected estimates of sensitivity and specificity.

DISCUSSION

The linear approach of the hierarchical assessment of diagnostic imaging technology is a reflection of the philosophy prevalent in the industrial period (Fig 1). A serial approach leads from technical development through various steps of technology assessment to implementation. In the postindustrial philosophy, an interwoven circular approach with concurrent development, assessment, and implementation of technology is more appropriate (Fig 7) (37,49). Initial technical development of new technology is followed by rapid implementation, and assessment begins concurrently. Assessment leads to adjustments in the number of requests for examinations and suggestions for refinement of the technology, which is the feedback mechanism that adjusts implementation and initiates technical development. This interwoven circular approach with feedback loops is typical of cybernetics and self-organizing systems (37).

The key features of our proposed study design (Table) are randomization, an empirically based pragmatic study protocol, an interweaving of research and clinical practice, outcome measures that focus on the clinical decision-making process, and measurement of trends over time in the outcome measures. These key features have both advantages and disadvantages that we will discuss.

Randomization for evaluating diagnostic tests is not novel, but, apart from screening trials, it has seldomly been

used for comparing diagnostic imaging modalities (39,50,51). Sometimes, randomization across diagnostic strategies in fact means randomization across management strategies that include both diagnostic work-up and treatment (52). This is clearly the case in screening trials in which not only the value of the diagnostic test but also the benefit of early diagnosis and treatment of the disease are being evaluated. The advantage of randomization clearly lies in the reduction of bias due to extraneous factors in the comparison between experimental and control groups.

Randomization also has disadvantages, however. The inherent limitations of randomization are reduced generalizability of study results due to the selected patient population and a perceived necessity to strictly control the experimental setting. Generalizability of trial results to other settings can be ensured by including a wide range of patients in the trial. Furthermore, by using an empirical pragmatic design we in fact study daily clinical practice, which increases generalizability.

Ethical considerations are frequently mentioned as an argument against randomization. If there are good reasons to believe that one diagnostic imaging modality is superior to another, then randomization is clearly unethical. In such situations, the inferior test should not be performed because of the possibility of misleading results, risks, and costs, and a trial is unnecessary. In fact, if the available evidence demonstrates that a new diagnostic strategy is superior, the results of the trial may even be misleading, or, at the very least, the research resources used are not justified relative to the scientific information obtained (35). Furthermore, if the tests are very different in their usage and goal (eg, D-dimer and spiral CT for help in diagnosing pulmonary embolism), then randomization between these two tests is neither useful nor feasible. Randomization between two full diagnostic strategies that include these tests, however, is feasible. In evaluating new, less invasive tests in the context of existing screening programs, randomization between the reference standard and the new test is ethical only if the available evidence suggests that the risk of missing the disease is more or less an equivalent impediment to the patient as are the risk and discomfort of performing the reference standard. For example, in screening patients with a history of premalignant colon polyps by using colonoscopy, randomization between colonoscopy and

virtual CT colonoscopy would be ethical only if the reduced burden to the patient of virtual colonoscopy is on the same order of magnitude as the burden to the patient of potentially missing a polyp. Finally, in the very early stages of development of a diagnostic technology and if missing the diagnosis would be life threatening, randomization between the new test and the reference standard is unacceptable.

More often than not, however, we truly do not know which diagnostic strategy is best, and diagnostic strategies have both potential advantages and disadvantages. Even obtaining all possible imaging information can have disadvantages, in that risks and burden are involved and results can potentially be misleading. In the example mentioned earlier, CT angiography may very well turn out to be diagnostically superior to intraarterial DSA, because multiple views can be obtained to evaluate the vessels. Furthermore, CT angiography is less risky and less burdensome than intraarterial DSA, which favors the use of the former. At the same time, however, physicians interpreting CT angiographic images may still be uncomfortable with the images, and artifacts may not be recognized as such. By allowing the use of an alternative diagnostic technique in cases where diagnostic uncertainty persists, valuable technology is not withheld from the patient. The premise is that physicians will request additional diagnostic examinations until they have sufficient diagnostic information to formulate a treatment plan.

Furthermore, the patient may be examined with the alternative test if symptoms persist or recur during follow-up. Finally, if we truly do not know which test strategy is better, randomization will at least ensure that 50% of patients will be assigned to the better strategy.

Costs are also frequently brought up as an argument against randomization. Randomization between two diagnostic tests, however, is less expensive than performing both tests in all patients. In fact, in the example of intraarterial DSA versus CT angiography, we claim that performing the trial may actually reduce the expense of the diagnostic imaging work-up for peripheral arterial disease in comparison with the expense of not performing the trial! If the trial had not been initiated, the work-up would continue to be performed with intraarterial DSA in all patients. Through randomization, half of all patients with peripheral arterial disease are initially evaluated with CT angiography, which, as the initial test, is clearly less expensive than intraarterial DSA. Depending on how often intraarterial DSA is performed after CT angiography, this can lead to a cost reduction. Furthermore, after CT angiography, an additional intraarterial DSA examination can often be limited to selective views, and an interventional procedure can usually be planned concurrently, which can lead to additional cost savings. The only real research expenses are those associated with data collection, data analysis, and reporting of the results—but that is a cost that will be incurred irrespective of

Key Features of Proposed Study Design for Concurrent Development, Assessment, and Implementation of New Imaging Technology	
Feature	Explanation
Randomization	Random assignment to either current diagnostic strategy (control) or new technology strategy (experimental); minimizes bias by extraneous factors
Empirical and pragmatic	Document effect on clinical decision making; feedback helps improve imaging protocol and postprocessing; feedback may influence recruitment rate
Self-organizing system	Subjective experience influences further technical development and adjustment of implementation
Integration	Integrate research and clinical practice in an interwoven process
Trend analysis	Outcome measures analyzed as function of time since start of study; demonstration of trends is a key result
Process-oriented outcome measures	Outcome measures should reflect clinical decision-making process and could include trends in (a) additional examinations requested, (b) costs of diagnostic work-up and therapy, (c) physician confidence in decision making, (d) recruitment rate
Patient outcomes	Include measures that reflect overall goal of health care: (a) quality of life, recurrent symptoms, events; (b) crossovers to other diagnostic strategy during follow-up

whether the trial is designed as a randomized or a nonrandomized trial. The proposed trial design will likely be funded by agencies interested in funding technology assessment, implementation, and translational research. We have been successful in obtaining funding from such an agency to perform the study described in this article.

An empirically based pragmatic study protocol may be frowned on by researchers—it refutes the notion that research should be performed in a strictly controlled setting. An empirical pragmatic study protocol may be influenced by subjective experience and individuals' opinions. Through feedback, we allow subjective experience to affect the postprocessing methods of the imaging study and the interpretation of images. Furthermore, we allow feedback to influence requests for additional examinations. The question is whether this empirical pragmatic approach invalidates the results of the trial. Our premise in allowing feedback to influence the clinical process under evaluation is that it will happen anyway—if not explicitly, it will happen implicitly. It is impossible to prevent an increase in experience and knowledge among technologists and physicians using the new techniques, and changes will take place, although sometimes in a subtle way. In general, a scientist cannot play the role of a detached objective observer but becomes involved in the world he or she observes (49).

Similarly, we can question whether we want to separate the clinical process and research. Rather than struggle to create a "pure" experimental setting, we question whether creating a pure experimental setting is desirable. One of the frequently expressed complaints about research is the delay between the reporting of results and the implementation thereof. Then why not integrate the clinical process and the research that is meant to improve that process? Why not make it into one interwoven whole? The clinical researcher is intricately involved in the clinical process, and, therefore, it would make immanent sense to integrate clinical research and clinical practice into one interwoven process (37).

Our chosen outcome measures are, in part, subjective and may therefore be prone to bias. Although this certainly represents a limitation of our approach, one should recognize that whatever we measure is determined and colored by our perception of the problem. For example, if we choose to measure the ankle-brachial blood pressure index (an objective outcome measure) to determine the severity and progression of peripheral arterial disease, we will observe hemodynamic changes due to obstruction in the blood vessels to the legs. The fact that the patient's symptoms may be influenced by the ability of muscle tissue to adapt to ischemia, the ability of vessels to develop collateral circulation, the patient's general fitness, and the patient's psychologic ability to cope with pain will remain unobserved. By using the ankle-brachial index as measure of disease severity, we have defined the problem of peripheral arterial disease as analogous to the problem encountered by a plumber. Thus, our choice of outcome measure is in itself subjective. We perceive the world from our own viewpoint, and our observations probably say more about how we have defined the problem and categorized the possible outcomes than about the actual phenomena we observe (49).

Ideally, one would want to observe phenomena from multiple perspectives. Putting all the different perspectives together will give an overall holistic picture. In choosing outcome measures for our trial design, we have attempted to capture the perspective of the patient by measuring health-related quality of life and events during follow-up, the perspective of the physician by measuring his or her confidence in the diagnosis and the decision whether or not to request further diagnostic examinations, and the perspective of the health care system by measuring the costs related to the diagnostic work-up and therapy. Depending on the clinical problem, it may be relevant to measure other outcomes as well.

Most important of all, we propose using time as an important dimension in the analysis of outcomes. Instead of measuring cumulative outcomes, the key feature of our proposed study design is to measure trends in outcomes over time. There is always a learning curve to get acquainted with new technology. Even after implementation of a new technology, adjustments are constantly made to the technique used, and interpretation skills keep developing. Thus, the development of new technology continues as we use it—and it should continue to develop! If we force ourselves to adhere to the exact same imaging protocol as the one with which we initially started, we will very quickly be using an outdated technique. Apart from changes in the technique used, it is impossible (and undesirable) to prevent interpreters from learning from their experience. Thus, instead of forcing the imaging technique and interpretation to remain stable and stationary, we propose using time as an explanatory variable in the data analyses, which will model the learning curve, technical developments, and increasing interpretation skills over time.

Our proposed approach is not a panacea: It has both advantages and limitations. As with every technique, model, and theory, this design will have to find its place among the multitude of study designs available. The approach seems especially useful when available evidence based on results from small clinical studies, exploratory decision- and cost-effectiveness analyses, and clinical experience suggests that two diagnostic strategies are similar, that both have advantages and disadvantages, that defining a reference standard proves to be difficult, or that the new technology could potentially surpass the existing reference standard. Furthermore, we would suggest the integration of structured data collection into daily clinical practice for the purpose of assessment of imaging technology and randomization between diagnostic strategies, rather than performance of new imaging tests in addition to the currently used tests. This enables assessment of new imaging technologies as they become available for clinical use.

In conclusion, for the development, assessment, and implementation of new diagnostic imaging technology, we propose use of a randomized empirical trial design based on a pragmatic study protocol that interweaves research and clinical practice. Outcome measures should include measures related to the clinical decision-making process, costs, and patient health outcomes. The key feature of our approach is to measure the trends over time in the outcome measures.

References

1. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. CMAJ 1986; 134:587–594.
2. Freedman LS. Evaluating and comparing imaging techniques: a review and classification of study designs. Br J Radiol 1987; 60:1071–1081.
3. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991; 11:88–94.
4. Kent DL, Larson EB. Disease, level of impact, and quality of research methods: three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. Invest Radiol 1992; 27:245–254.
5. Taylor CR, Elmore JG, Sun K, Inouye SK. Technology assessment in diagnostic imaging: a proposal for a phased approach to evaluating radiology research. Invest Radiol 1993; 28:155–161.
6. Thornbury JR. Eugene W. Caldwell Lec-

ture: clinical efficacy of diagnostic imaging—love it or leave it. AJR Am J Roentgenol 1994; 162:1–8.

7. van der Schouw YT, Verbeek AL, Ruijs SH. Guidelines for the assessment of new diagnostic tests. Invest Radiol 1995; 30: 334–340.

8. Mackenzie R, Dixon AK. Measuring the effects of imaging: an evaluative framework. Clin Radiol 1995; 50:513–518.

9. Hunink MGM. Cost-effectiveness research in radiology. Acad Radiol 1996; 3(suppl 1):S13–S16.

10. Hunink MG. Outcomes research and cost-effectiveness analysis in radiology. Eur Radiol 1996; 6:615–620.

11. Hillman BJ. Outcomes research in radiology. Acad Radiol 1996; 3(suppl 1):S3–S4.

12. Pearl WS. A hierarchical outcomes approach to test assessment. Ann Emerg Med 1999; 33:77–84.

13. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York, NY: Oxford University Press, 1996.

14. Goldratt EM. The goal: a process of ongoing improvement. Great Barrington, Mass: North River Press, 1992.

15. Goldratt EM. It's not luck. Great Barrington, Mass: North River Press, 1995.

16. Baum RA, Rutter CM, Sunshine JH, et al. Multicenter trial to evaluate vascular magnetic resonance angiography of the lower extremity. JAMA 1995; 274:875–880.

17. Yin D, Baum RA, Carpenter JP, Langlotz CP, Pentecost MJ. Cost-effectiveness of MR angiography in cases of limb-threatening peripheral vascular disease. Radiology 1995; 194:757–764.

18. Rieker O, Duber C, Schmiedt W, von Zitzewitz H, Schweden F, Thelen M. Prospective comparison of CT angiography of the legs with intraarterial digital subtraction angiography. AJR Am J Roentgenol 1996; 166:269–276.

19. Rieker O, Duber C, Neufang A, Pitton M, Schweden F, Thelen M. CT angiography versus intraarterial digital subtraction angiography for assessment of aortoiliac occlusive disease. AJR Am J Roentgenol 1997; 169:1133–1138.

20. Ruehm SG, Hany TF, Pfammatter T, Schneider E, Ladd M, Debatin JF. Pelvic and lower extremity arterial imaging: diagnostic performance of three-dimensional contrast-enhanced MR angiography. AJR Am J Roentgenol 2000; 174:1127–1135.

21. Begg CB, McNeil BJ. Assessment of radiologic tests: control of bias and other design considerations. Radiology 1988; 167: 565–569.

22. Begg CB. Biases in the assessment of diagnostic tests. Stat Med 1987; 6:411–423.

23. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282:1061–1066.

24. Weller SC, Mann NC. Assessing rater performance without a "gold standard" using consensus theory. Med Decis Making 1997; 17:71–79.

25. Phelps CE, Hutson A. Estimating diagnostic test accuracy using a "fuzzy gold standard." Med Decis Making 1995; 15:44–57.

26. Begg CB, Greenes A. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983; 39:207–215.

27. Begg CB. Estimation of risks when verification of disease status is obtained in a selected group of subjects. Am J Epidemiol 1984; 120:328–330.

28. Hunink MGM, Richardson D, Doubilet PM, Begg CB. Testing for fetal pulmonary maturity: an ROC analysis involving covariates, verification bias and combination testing. Med Decis Making 1990; 10: 201–211.

29. Hunink MGM, Polak JF, Barlan MM, O'Leary DH. Detection and quantification of carotid artery stenosis: efficacy of various Doppler velocity parameters. AJR Am J Roentgenol 1993; 160:619–625.

30. Lijmer JG, Hunink MG, van den Dungen JJ, Loonstra J, Smit AJ. ROC analysis of noninvasive tests for peripheral arterial disease. Ultrasound Med Biol 1996; 22: 391–398.

31. Hunink MGM, Glasziou PP, Siegel JE, et al. Decision making in health and medicine: integrating evidence and values. Cambridge, England: Cambridge University Press, 2001.

32. Kent KC, Kuntz KM, Patel MR, et al. Perioperative imaging strategies for carotid endarterectomy: an analysis of morbidity and cost-effectiveness in symptomatic patients. JAMA 1995; 274:888–893.

33. Kuntz KM, Fleischmann KE, Hunink MG, Douglas PS. Cost-effectiveness of diagnostic strategies for patients with chest pain. Ann Intern Med 1999; 130:709–718.

34. Wright JC, Weinstein MC. Gains in life expectancy from medical interventions: standardizing data on outcomes. N Engl J Med 1998; 339:380–386.

35. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ 1999; 18:341–364.

36. Eisenberg JM. Ten lessons for evidence-based technology assessment. JAMA 1999; 282:1865–1869.

37. Capra F. The web of life. New York, NY: Doubleday, 1996.

38. Pocock SJ. Clinical trials: a practical approach. Chichester, England: Wiley, 1983.

39. van Loon AJ, Mantingh A, Serlier EK, Kroon G, Mooyaart EL, Huisjes HJ. Randomised controlled trial of magnetic-resonance pelvimetry in breech presentation at term. Lancet 1997; 350:1799–1804.

40. Dixon AK. Clinical value of computed tomography scanning and magnetic resonance imaging: studies in the United Kingdom. Acad Radiol 1996; 3:S49–S50.

41. Dixon AK. The impact of medical imaging on the physician's diagnostic and therapeutic thinking. Eur Radiol 1998; 8:488–490.

42. Jarvik JG. Health profiles in quality of life: evaluation of diagnostic technologies. Acad Radiol 1999; 6(suppl 1):S93–S97.

43. Bosch JL, van der Graaf Y, Hunink MG. Health-related quality of life after angioplasty and stent placement in patients with iliac artery occlusive disease: results of a randomized controlled clinical trial. Circulation 1999; 99:3155–3160.

44. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. JAMA 1996; 276:637–639.

45. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36). II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993; 31:247–263.

46. Ware JE Jr, Sherbourne CD. The MOS 36-Item Short-form Health Survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30:473–483.

47. Morgan MB, Crayford T, Murrin B, Fraser SC. Developing the Vascular Quality of Life Questionnaire: a new disease-specific quality of life measure for use in lower limb ischemia. J Vasc Surg 2001; 33:679–687.

48. Brooks R. EuroQol: the current state of play. Health Policy 1996; 37:53–72.

49. Capra F. The tao of physics. Boston, Mass: Shambhala, 1975.

50. Jarvik JG, Deyo RA, Koepsell TD. Screening magnetic resonance images versus plain films for low back pain: a randomized trial of effects on patient outcomes. Acad Radiol 1996; 3(suppl 1):S28–S31.

51. Jarvik JG, Maravilla KR, Haynor DR, Levitz M, Deyo RA. Rapid MR imaging versus plain radiography in patients with low back pain: initial results of a randomized study. Radiology 1997; 204:447–454.

52. Bossuyt PM, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. Lancet 2000; 356:1844–1847.