Different forms of evidence for different types of questions

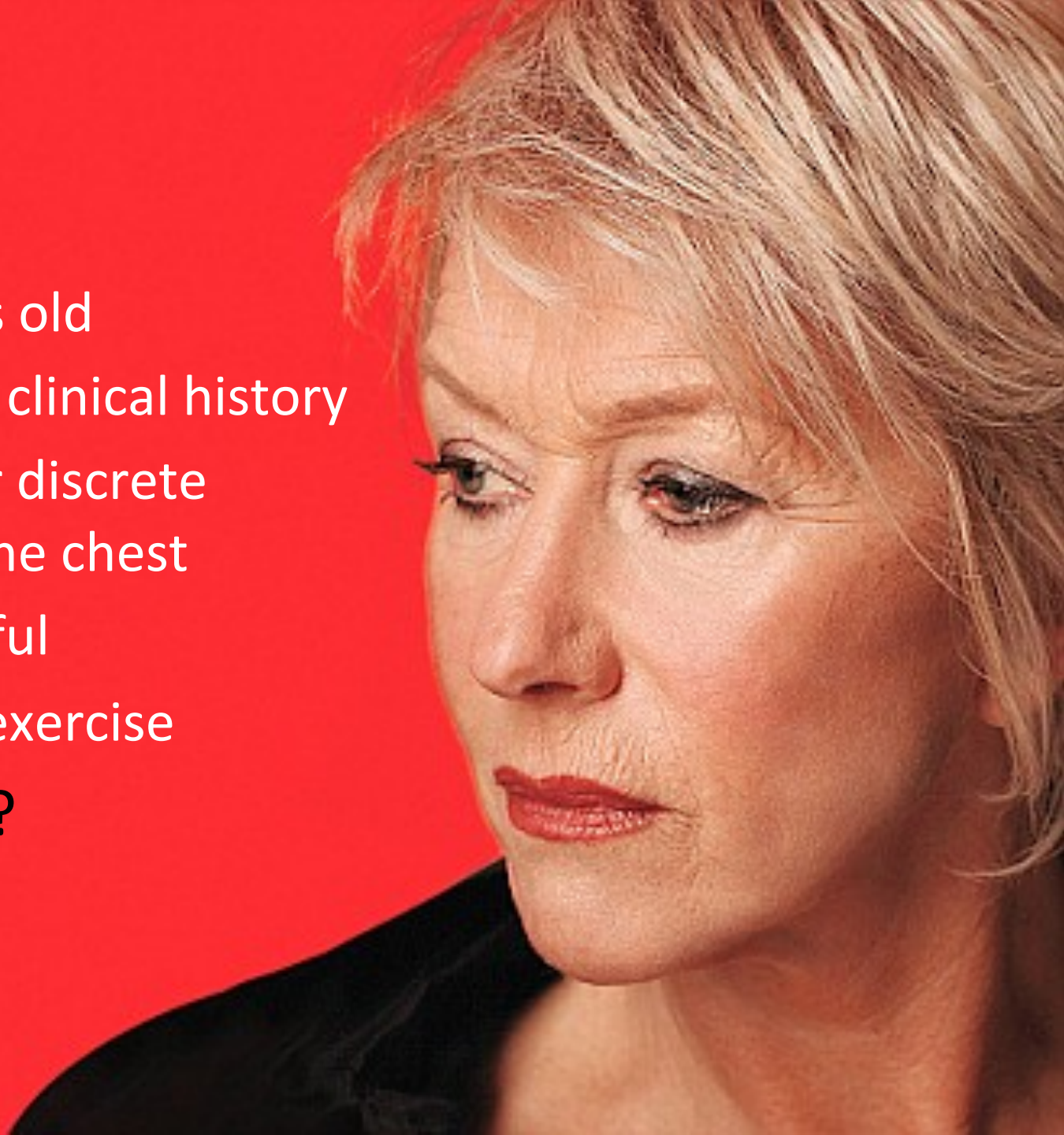
Ann Van den Bruel

Director Diagnostic Evidence Co-operative
Senior Clinical Research Fellow
Department of Primary Care Health Sciences

This session

- How to frame a diagnostic question
- Design diagnostic accuracy studies
- Design impact studies

- Helen, 67 years old
- No remarkable clinical history
- Sees her GP for discrete discomfort in the chest
- Not really painful
- Worsens with exercise
- Exercise ECG?



Formulate your question

- P
- I
- C
- O

- P
- I
- R
- T

PIRT

Let's exercise

- P — How would I describe a group of patients similar to mine?
- I — Which test am I considering?
- R — What is the reference standard considered to be ideal to diagnose the target condition?
- T — Which target condition/diagnosis do I want to either rule in or rule out?

melanoma patients and to determine the diagnostic value of subsequent PET/CT and MRI of the brain in these patients.

PATIENTS AND METHODS

Patients

Between August 2006 and March 2009, 46 melanoma patients without symptoms and signs of recurrent disease were referred for total body PET/CT and MRI of the brain because of an increased S-100B. The mean age of the patients was 59 years (range 25–93 years). Serum S-100B was monitored during follow-up after the surgical treatment of regional or distant metastases or because a patient was at increased risk due to primary tumor features (Table 1).

S-100B Analysis

The S-100B concentration was determined in serum using the Elecsys S100 assay, which is an electrochemiluminescence immunoassay (ECLIA) for the in vitro quantitative determination of S100 (S100 A1B and S100 BB) in human serum (Roche Diagnostics, Mannheim, Germany). The immunoassay ECLIA is intended for use on Elecsys and cobas e immunoassay analyzers as described in detail previously.[12] In our laboratory, the upper reference value of S-100B has been established at 0.10 μg/L. In cases of an increased S-100B level, sampling and measurement of the tumor marker were repeated for confirmation within a few days. Only patients in whom the repeat value was also increased were enrolled in the study.

FDG PET/CT

A hybrid PET/CT camera (Gemini II, Philips, Eindhoven, The Netherlands) was used, and FDG was administrated in dosages of 180–240 MBq (4.9–6.5 mCi). PET/CT scans were performed after fasting for 6 hours. The interval between FDG administration and scanning was 60 minutes ± 10 minutes. Low-dose CT images (40 mAs, 5 mm slices) were acquired without oral or intravenous contrast. Generated images were displayed using an Osirix Dicom viewer in a Unix-based operating system (MAC OS X, Power G5, Apple, Cupertino, CA) and were evaluated on the basis of two-dimensional orthogonal reslicing. PET was fused to low-dose CT after correction for attenuation. The PET/CT scans were reviewed by 3 experienced nuclear medicine physicians together.

MRI

MRI was performed with a high-field strength 3.0 T scanner (Achieva, Philips, Eindhoven, The Netherlands). The protocol consisted of precontrast transversal T2-weighted imaging, axial fluid attenuated inversion recovery (FLAIR) imaging, diffusion-weighted imaging and precontrast and postcontrast coronal T1-weighted 3D-FFE imaging.

Reference Standard

The presence or absence of melanoma recurrence was established by fine needle aspiration cytology or histological biopsy when possible. Additional imaging and the clinical course were used as the gold standard if no pathologic result could be obtained.

Statistical Analysis

Statistical analyses were performed using SPSS 15 (Version 15, for Windows, SPSS Inc, Chicago, IL). The accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of PET/CT for the detection of local-regional recurrence or distant metastases were calculated using the standard definitions. Kaplan-Meier curves were used to analyze survival and were compared using a two-sided log-rank test. A difference was considered statistically significant if the associated *P* value was .05 or less.


Think
…..about a diagnostic question
SCIENCEphotoLIBRARY

Think

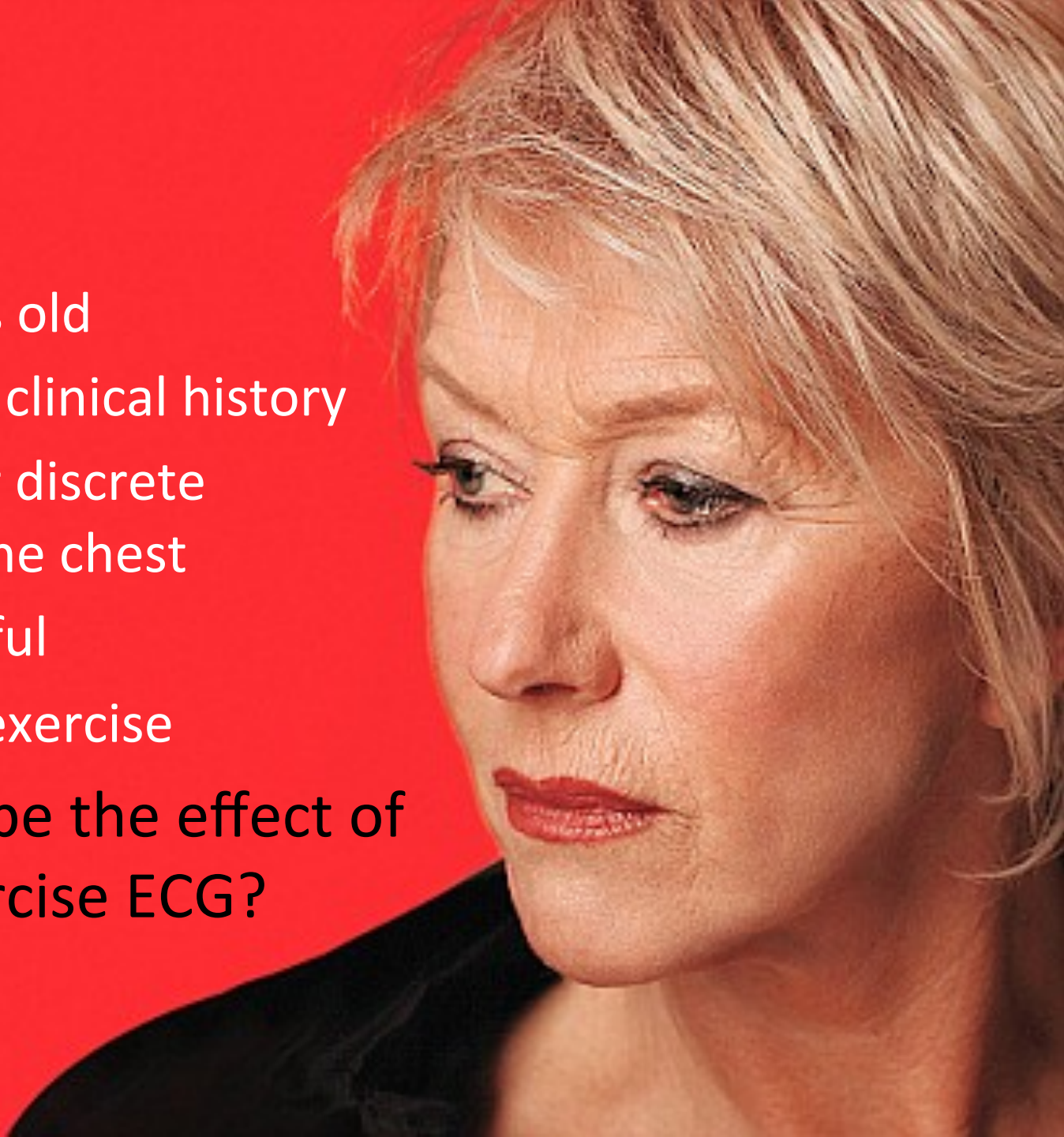
- What's your PIRT?

PIRT

- P — How would I describe a group of patients similar to mine?
- I — Which test am I considering?
- R — What is the reference standard considered to be ideal to diagnose the target condition?
- T — Which target condition/diagnosis do I want to either rule in or rule out?

Create your own diagnostic accuracy study

- Helen, 67 years old
- No remarkable clinical history
- Sees her GP for discrete discomfort in the chest
- Not really painful
- Worsens with exercise
- What would be the effect of doing an exercise ECG?



Formulate your question

- P
- I
- C
- O

- P
- I
- R
- T

PICO

Let's exercise

- P — How would I describe a group of patients similar to mine?
- I — Which test am I considering?
- C — What is the diagnostic strategy I would like to compare with?
- O — What are the outcomes that the new test could affect?

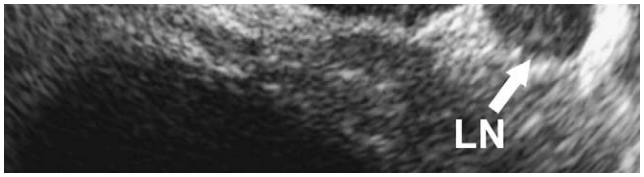


Figure 4. 270° transverse-array EUS image of the esophageal malignancy shown in Fig. 1, staged T2 N1. An adjacent malignant lymph node is illustrated by *LN*.

the 2 modalities in the staging of cancer of the esophagus or the cardia.[2]

In this study, we aimed to compare the new method of electronic 270° transverse-array EUS with L-EUS for the staging of upper-GI (UGI) malignancies.

MATERIALS AND METHODS

Institutional review board approval was obtained for 50 patients to be enrolled for this study. The staging by a linear array endosonoscope (EG-3630U; Pentax America Inc, Montvale, NJ) (Fig. 2) was compared with staging by a recently introduced electronic 270° transverse-array endosonoscope (EG-3630UR; Pentax) (Fig. 3). The staging consisted of tumor (T) and nodal (N) staging. T staging consisted of determining the depth of invasion, as well as measurements of the thickness of the tumor. N staging was performed by determining the number and the size of enlarged lymph nodes adjacent to the tumor and in distant nodal sites.

Patients undergoing staging of GI malignancy with EUS were enrolled in the study. Patients with an obstructed esophagus that could not be passed with an endoscope were excluded from the study. In addition, uncooperative patients or subjects without informed consent were not enrolled. Subjects were recruited from the inpatient and outpatient practices of the investigators without the use of advertisement.

Subjects underwent endoscopy and EUS with standard instruments. This study involved the comparison of 2 EUS examinations, L-EUS and transverse-array EUS. The size, the location, and the stage of the tumor and the lymph nodes with each instrument were documented at each examination. After the examination, subjects were monitored for procedure-related complications, such as bleeding, perforation, and pain. The 2 endosonographers who conducted this study are experienced gastroenterologists with 8 and 12 years of experience in diagnostic and interventional EUS.

Specific data variables that were collected included location of the tumor, diameter and thickness of the tumor mass, depth of invasion into the organ wall, size and number of abnormal lymph nodes. The study was designed to detect a difference between the accuracy of tumor staging by 2 different instruments. We hypothesized that the size and the number of abnormal lymph nodes would be greater with a transverse-array echoendoscope, compared with the linear array instrument.

Both EUS techniques were compared subjectively by determining an assessment score for image quality, clinical tumor-staging quality, and ease of intubation on an ordinal scale of 1 to 5 (1, lowest; 5, highest rating).

Definitions

For the staging of lymph nodes, we defined N0 as a nonmalignant-appearing lymph node (lymph node size <10 mm, isoechoic texture) visible by EUS or a suspicious lymph node with nonmalignant cytology.[3,4] We defined N1 as a malignant-appearing lymph node with a cytology that demonstrated malignancy (Table 1).

Statistical methods

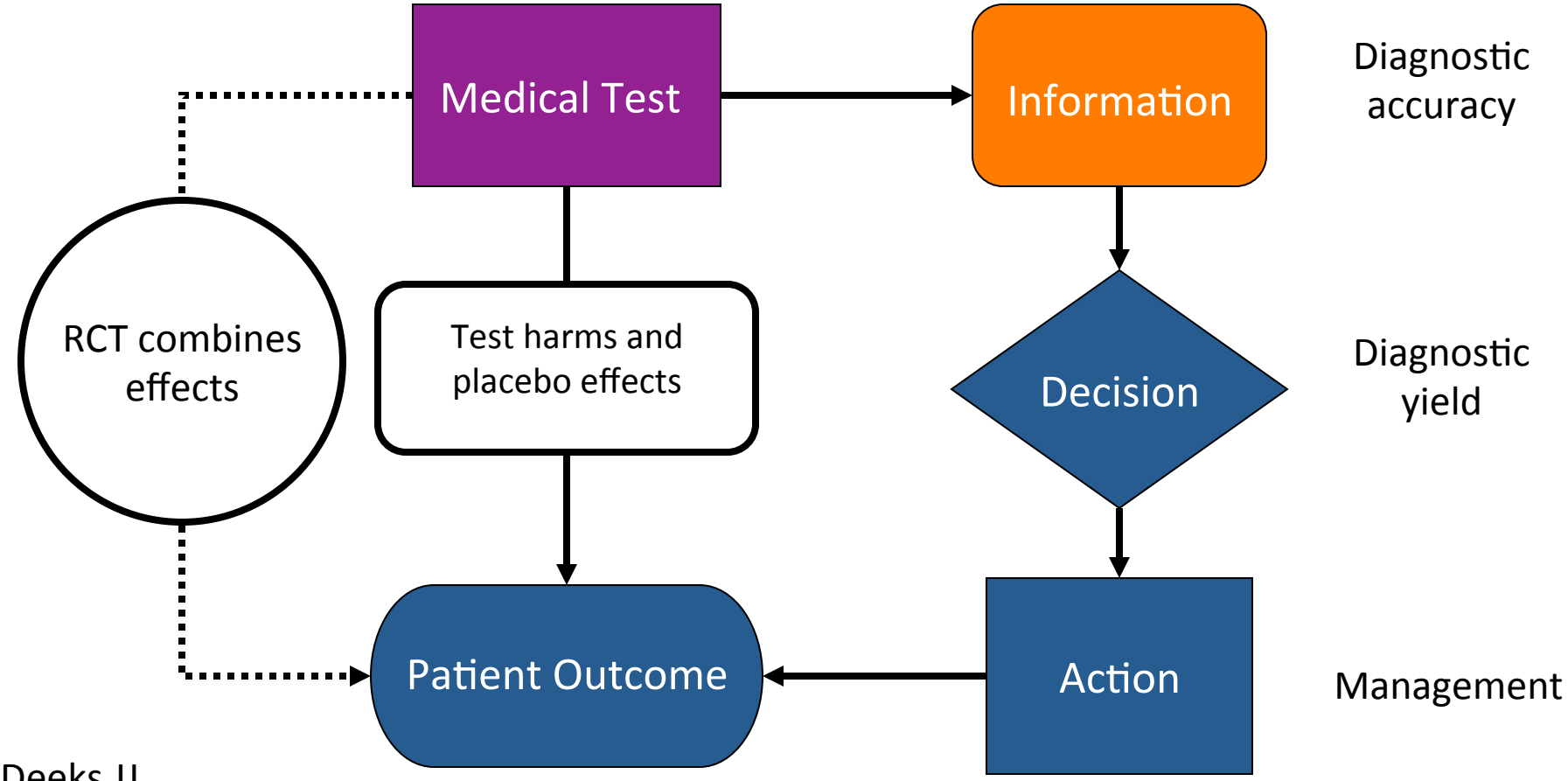
A Student t test or a χ^2 analysis, with a Yates correction for continuity where appropriate, was used to compare tumor and lymph-node staging by the 2 different types of echoendoscopes. A sample-size calculation was performed during the design of the study. By assuming that the average number of abnormal lymph nodes would be 2 for the traditional examination and an average of 3 abnormal lymph nodes would be identified with the transverse-array echoendoscope, approximately 50 examinations would be required to detect a significant difference. The medians of the assessment scores of both techniques were compared by using the Wilcoxon signed rank test. Results were considered as statistically significant if the P value was $<.05$.

Indications for diagnostic trials

- Tests detect disease earlier (screening and case-finding)
- Test itself has a harmful effect
- Interventions have harmful effects
  - Treating some non-diseased may outweigh benefits of treating diseased
- No reference standard

- Rare goods:
  - Only 37 (95% CI 35-40) diagnostic test strategies RCTs on patient outcomes per year.
  - 21,949 per year for all RCTs indexed in CENTRAL.

What is being evaluated?



Deeks JJ

Studying impact of tests

- On
  - Patient outcome
  - Costs
  - Organisation of care
- Designs:
  - RCT
  - Before-after trial
  - Modelling

What is being evaluated?

Trial finds no difference:

????

- **Conditions for a test to be of diagnostic benefit**
  - Test is more accurate
  - Interpretation of test results is rational and consistent
  - Management is rational and consistent
  - Treatment is effective

- **Conditions for a trial to be informative**
  - Rules for interpretation of test results are described
  - Management protocol is described

- **No descriptions given in example trials**
  - Applying the results requires faith that the behaviour of your patients and clinicians is the same as the trial

Clinically important differences

Let's exercise

Test delivery	Timing of test
	Feasibility
	Test process
Test result	Interpretability
	Accuracy
	Timing of results
Diagnostic decision	Timing of diagnosis
	Diagnostic confidence
Treatment decision	Therapeutic yield
	Therapeutic confidence
Treatment implementation	Time to treatment
	Efficacy of treatment
	Adherence to treatment

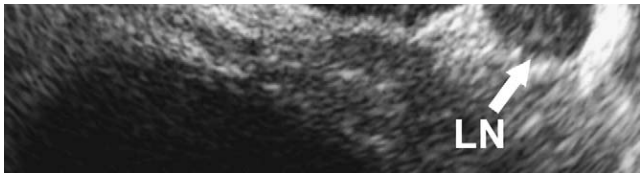


Figure 4. 270° transverse-array EUS image of the esophageal malignancy shown in Fig. 1, staged T2 N1. An adjacent malignant lymph node is illustrated by *LN*.

the 2 modalities in the staging of cancer of the esophagus or the cardia.[2]

In this study, we aimed to compare the new method of electronic 270° transverse-array EUS with L-EUS for the staging of upper-GI (UGI) malignancies.

MATERIALS AND METHODS

Institutional review board approval was obtained for 50 patients to be enrolled for this study. The staging by a linear array endosonoscope (EG-3630U; Pentax America Inc, Montvale, NJ) (Fig. 2) was compared with staging by a recently introduced electronic 270° transverse-array endosonoscope (EG-3630UR; Pentax) (Fig. 3). The staging consisted of tumor (T) and nodal (N) staging. T staging consisted of determining the depth of invasion, as well as measurements of the thickness of the tumor. N staging was performed by determining the number and the size of enlarged lymph nodes adjacent to the tumor and in distant nodal sites.

Patients undergoing staging of GI malignancy with EUS were enrolled in the study. Patients with an obstructed esophagus that could not be passed with an endoscope were excluded from the study. In addition, uncooperative patients or subjects without informed consent were not enrolled. Subjects were recruited from the inpatient and outpatient practices of the investigators without the use of advertisement.

Subjects underwent endoscopy and EUS with standard instruments. This study involved the comparison of 2 EUS examinations, L-EUS and transverse-array EUS. The size, the location, and the stage of the tumor and the lymph nodes with each instrument were documented at each examination. After the examination, subjects were monitored for procedure-related complications, such as bleeding, perforation, and pain. The 2 endosonographers who conducted this study are experienced gastroenterologists with 8 and 12 years of experience in diagnostic and interventional EUS.

Specific data variables that were collected included location of the tumor, diameter and thickness of the tumor mass, depth of invasion into the organ wall, size and number of abnormal lymph nodes. The study was designed to detect a difference between the accuracy of tumor staging by 2 different instruments. We hypothesized that the size and the number of abnormal lymph nodes would be greater with a transverse-array echoendoscope, compared with the linear array instrument.

Both EUS techniques were compared subjectively by determining an assessment score for image quality, clinical tumor-staging quality, and ease of intubation on an ordinal scale of 1 to 5 (1, lowest; 5, highest rating).

Definitions

For the staging of lymph nodes, we defined N0 as a nonmalignant-appearing lymph node (lymph node size <10 mm, isoechoic texture) visible by EUS or a suspicious lymph node with nonmalignant cytology.[3,4] We defined N1 as a malignant-appearing lymph node with a cytology that demonstrated malignancy (Table 1).

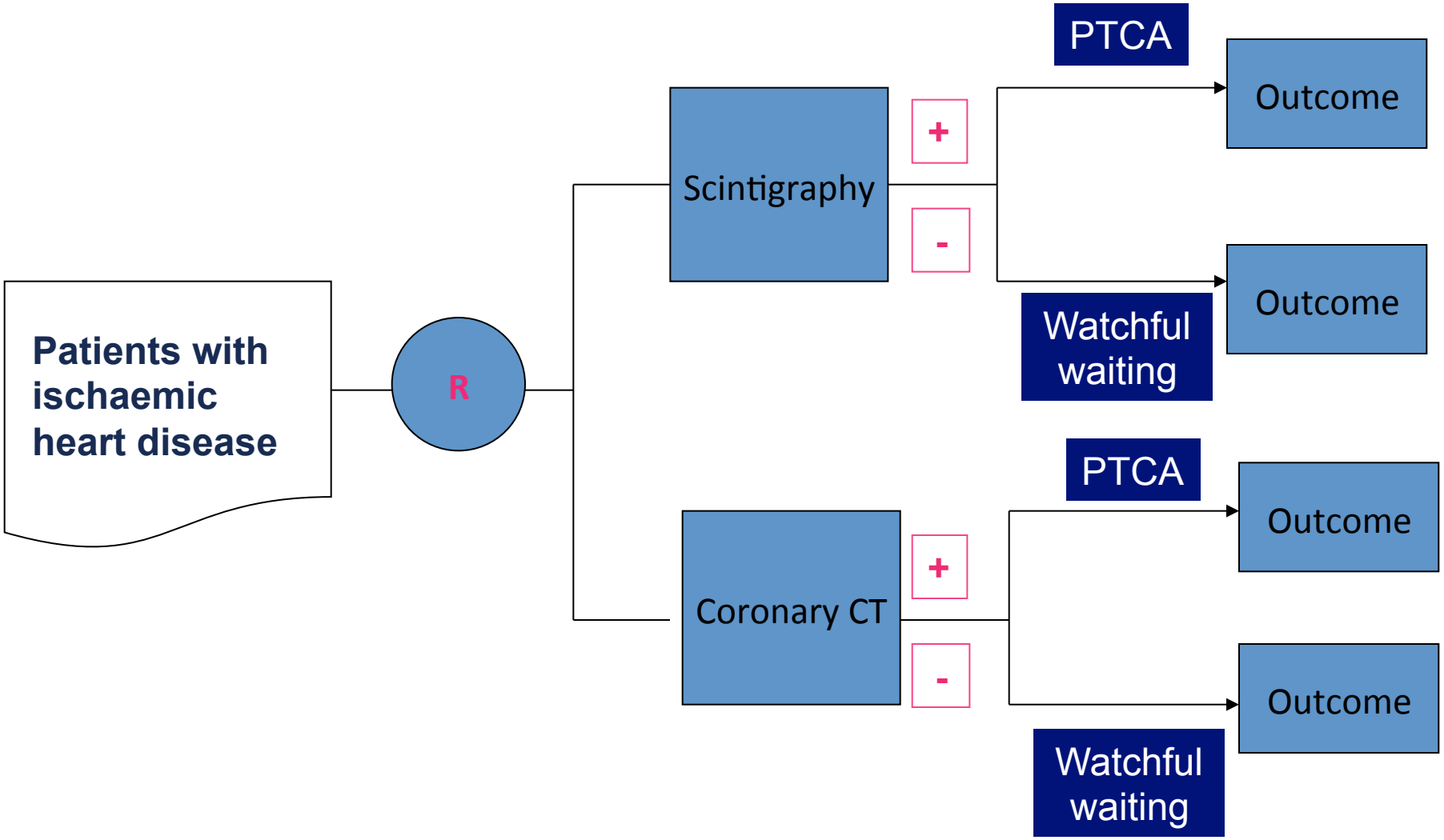
Statistical methods

A Student t test or a χ^2 analysis, with a Yates correction for continuity where appropriate, was used to compare tumor and lymph-node staging by the 2 different types of echoendoscopes. A sample-size calculation was performed during the design of the study. By assuming that the average number of abnormal lymph nodes would be 2 for the traditional examination and an average of 3 abnormal lymph nodes would be identified with the transverse-array echoendoscope, approximately 50 examinations would be required to detect a significant difference. The medians of the assessment scores of both techniques were compared by using the Wilcoxon signed rank test. Results were considered as statistically significant if the P value was $<.05$.

Imagine… direct impact?

Effects of testing	What this means	Effects on health
Emotional	Test causes harmful or beneficial changes in levels of anxiety, depression, stress, psychological well being.	Increased anxiety and stress occurring following a positive test on screening that has not been confirmed with a reference standard. Reassurance and improved overall well-being after a negative test.
Social	Effects of testing on social roles, social functions, sexual relationships, social relationships.	Social isolation and stigmatisation after a positive test. Problems with employment or insurance coverage. Genetic testing results may cause guilt about passing on a genetic predisposition.
Cognitive	Patients' beliefs, perceptions and understanding about the test result and the condition.	May understand disease better – what causes it, how long it lasts etc., or affect adherence to therapy.
Behavioural	The combinations of emotional, social and cognitive effects can affect patient behaviour. Positive and negative tests can prompt change in behaviour.	Adherence to clinical intervention may be increased or decreased. Greater or less engagement with other health related behaviours, e.g. increased exercise after having cholesterol measured. Perceptions of risks from screening and repeated screening.

RCT architecture



Journal of the American College of Cardiology
© 2011 by the American College of Cardiology Foundation
Published by Elsevier Inc.

Vol. 58, No. 14, 2011
ISSN 0735-1097/$36.00
doi:10.1016/j.jacc.2011.03.068

CLINICAL RESEARCH | Clinical Trial

The CT-STAT (Coronary Computed Tomographic Angiography for Systematic Triage of Acute Chest Pain Patients to Treatment) Trial

James A. Goldstein, MD,* Kavitha M. Chinnaiyan, MD,* Aiden Abidov, MD, PHD,† Stephan Achenbach, MD,‡ Daniel S. Berman, MD,§ Sean W. Hayes, MD,§ Udo Hoffmann, MD,‖ John R. Lesser, MD,¶ Issam A. Mikati, MD,# Brian J. O'Neil, MD,** Leslee J. Shaw, PHD,†† Michael Y. H. Shen, MD,‡‡ Uma S. Valeti, MBBS,§§ Gilbert L. Raff, MD,*
for the CT-STAT Investigators

Royal Oak and Detroit, Michigan; Tucson, Arizona; Giessen, Germany; Los Angeles, California; Boston, Massachusetts; Minneapolis, Minnesota; Chicago, Illinois; Atlanta, Georgia; and Fort Lauderdale, Florida

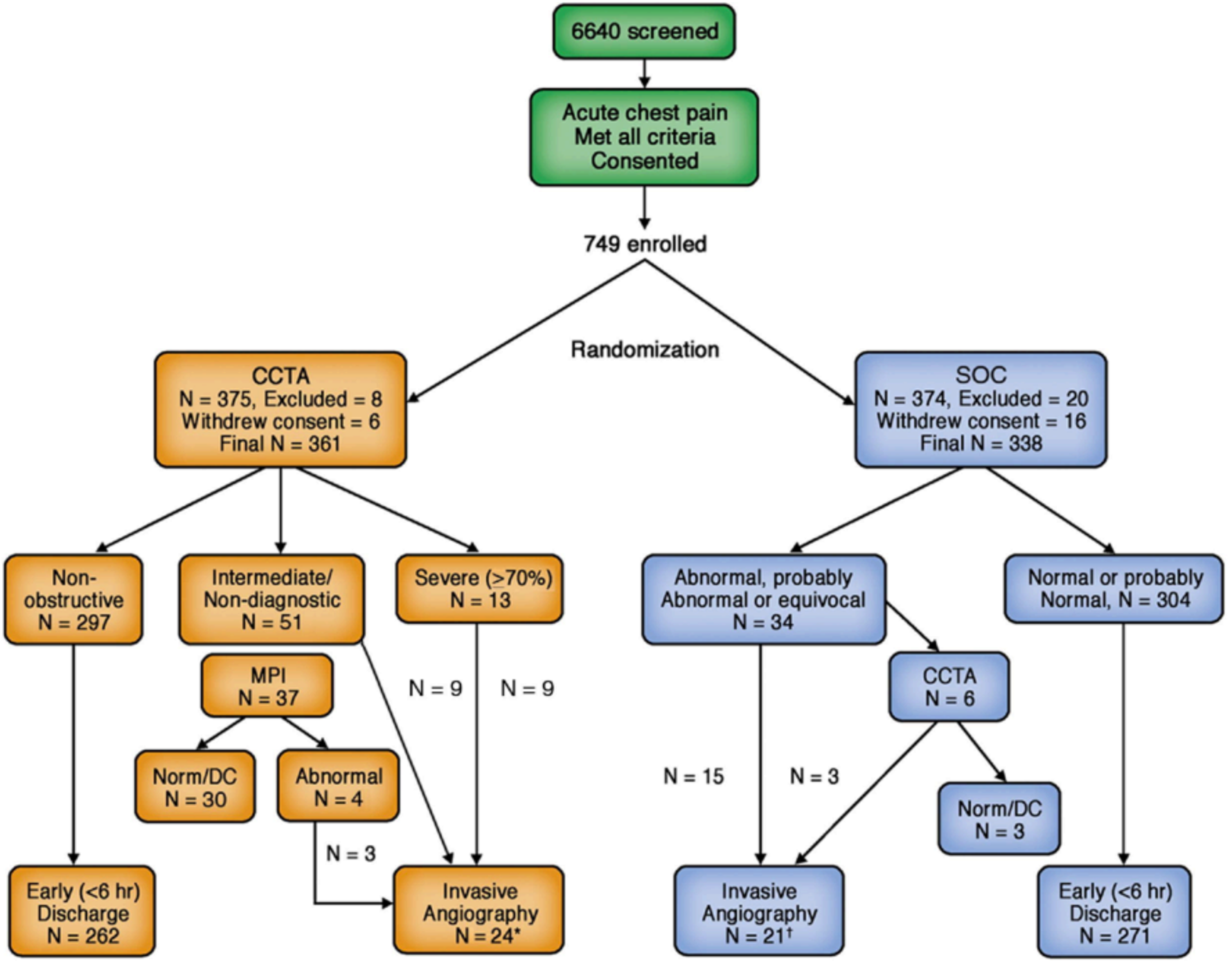
Objectives The purpose of this study was to compare the efficiency, cost, and safety of a diagnostic strategy employing early coronary computed tomographic angiography (CCTA) to a strategy employing rest-stress myocardial perfusion imaging (MPI) in the evaluation of acute low-risk chest pain.

Background In the United States, >8 million patients require emergency department evaluation for acute chest pain annually at an estimated diagnostic cost of >$10 billion.

Methods This multicenter, randomized clinical trial in 16 emergency departments ran between June 2007 and November 2008. Patients were randomly allocated to CCTA (n = 361) or MPI (n = 338) as the index noninvasive test. The primary outcome was time to diagnosis; the secondary outcomes were emergency department costs of care and safety, defined as freedom from major adverse cardiac events in patients with normal index tests, including 6-month follow-up.

Results The CCTA resulted in a 54% reduction in time to diagnosis compared with MPI (median 2.9 h [25th to 75th percentile: 2.1 to 4.0 h] vs. 6.3 h [25th to 75th percentile: 4.2 to 19.0 h], $p < 0.0001$). Costs of care were 38% lower compared with standard (median $2,137 [25th to 75th percentile: $1,660 to $3,077] vs. $3,458 [25th to 75th percentile: $2,900 to $4,297], $p < 0.0001$). The diagnostic strategies had no difference in major adverse cardiac events after normal index testing (0.8% in the CCTA arm vs. 0.4% in the MPI arm, $p = 0.29$).

Conclusions In emergency department acute, low-risk chest pain patients, the use of CCTA results in more rapid and cost-efficient safe diagnosis than rest-stress MPI. Further studies comparing CCTA to other diagnostic strategies are needed to optimize evaluation of specific patient subsets. (Coronary Computed Tomographic Angiography for Systematic Triage of Acute Chest Pain Patients to Treatment [CT-STAT]; NCT00468325) (J Am Coll Cardiol 2011;58:1414–22) © 2011 by the American College of Cardiology Foundation


6640 screened
Acute chest pain
Met all criteria
Consented
749 enrolled
Randomization
CCTA
N = 375, Excluded = 8
Withdrew consent = 6
Final N = 361
SOC
N = 374, Excluded = 20
Withdrew consent = 16
Final N = 338
Non-obstructive
N = 297
Intermediate/
Non-diagnostic
N = 51
Severe (≥70%)
N = 13
MPI
N = 37
Norm/DC
N = 30
Abnormal
N = 4
N = 3
N = 9
N = 9
Early (<6 hr)
Discharge
N = 262
Invasive
Angiography
N = 24*
Abnormal, probably
Abnormal or equivocal
N = 34
Normal or probably
Normal, N = 304
CCTA
N = 6
N = 15
N = 3
Norm/DC
N = 3
Invasive
Angiography
N = 21†
Early (<6 hr)
Discharge
N = 271

Journal of the American College of Cardiology
© 2011 by the American College of Cardiology Foundation
Published by Elsevier Inc.

Vol. 58, No. 14, 2011
ISSN 0735-1097/$36.00
doi:10.1016/j.jacc.2011.03.068

CLINICAL RESEARCH | Clinical Trial

The CT-STAT (Coronary Computed Tomographic Angiography for Systematic Triage of Acute Chest Pain Patients to Treatment) Trial

James A. Goldstein, MD,* Kavitha M. Chinnaiyan, MD,* Aiden Abidov, MD, PHD,† Stephan Achenbach, MD,‡ Daniel S. Berman, MD,§ Sean W. Hayes, MD,§ Udo Hoffmann, MD,‖ John R. Lesser, MD,¶ Issam A. Mikati, MD,# Brian J. O'Neil, MD,** Leslee J. Shaw, PHD,†† Michael Y. H. Shen, MD,‡‡ Uma S. Valeti, MBBS,§§ Gilbert L. Raff, MD,*
for the CT-STAT Investigators

Royal Oak and Detroit, Michigan; Tucson, Arizona; Giessen, Germany; Los Angeles, California; Boston, Massachusetts; Minneapolis, Minnesota; Chicago, Illinois; Atlanta, Georgia; and Fort Lauderdale, Florida

Objectives The purpose of this study was to compare the efficiency, cost, and safety of a diagnostic strategy employing early coronary computed tomographic angiography (CCTA) to a strategy employing rest-stress myocardial perfusion imaging (MPI) in the evaluation of acute low-risk chest pain.

Background In the United States, >8 million patients require emergency department evaluation for acute chest pain annually at an estimated diagnostic cost of >$10 billion.

Methods This multicenter, randomized clinical trial in 16 emergency departments ran between June 2007 and November 2008. Patients were randomly allocated to CCTA (n = 361) or MPI (n = 338) as the index noninvasive test. The primary outcome was time to diagnosis; the secondary outcomes were emergency department costs of care and safety, defined as freedom from major adverse cardiac events in patients with normal index tests, including 6-month follow-up.

Results The CCTA resulted in a 54% reduction in time to diagnosis compared with MPI (median 2.9 h [25th to 75th percentile: 2.1 to 4.0 h] vs. 6.3 h [25th to 75th percentile: 4.2 to 19.0 h], $p < 0.0001$). Costs of care were 38% lower compared with standard (median $2,137 [25th to 75th percentile: $1,660 to $3,077] vs. $3,458 [25th to 75th percentile: $2,900 to $4,297], $p < 0.0001$). The diagnostic strategies had no difference in major adverse cardiac events after normal index testing (0.8% in the CCTA arm vs. 0.4% in the MPI arm, $p = 0.29$).

Conclusions In emergency department acute, low-risk chest pain patients, the use of CCTA results in more rapid and cost-efficient safe diagnosis than rest-stress MPI. Further studies comparing CCTA to other diagnostic strategies are needed to optimize evaluation of specific patient subsets. (Coronary Computed Tomographic Angiography for Systematic Triage of Acute Chest Pain Patients to Treatment [CT-STAT]; NCT00468325) (J Am Coll Cardiol 2011;58:1414–22) © 2011 by the American College of Cardiology Foundation

ORIGINAL INVESTIGATION

Cost-effectiveness of B-Type Natriuretic Peptide Testing in Patients With Acute Dyspnea

Christian Mueller, MD; Kirsten Laule-Kilian, BSc; Christian Schindler, PhD; Theresia Klima, MD; Barbara Frana, MD; Daniel Rodriguez, MD; André Scholer, PhD; Michael Christ, MD; André P. Perruchoud, MD

Background: B-type natriuretic peptide (BNP) is a quantitative marker of heart failure that seems to be helpful in its diagnosis.

Methods: We performed a prospective randomized study (B-Type Natriuretic Peptide for Acute Shortness of Breath Evaluation) including 452 patients who presented to the emergency department with acute dyspnea to estimate the long-term cost-effectiveness of BNP guidance. Participants were randomly assigned to a diagnostic strategy involving the measurement of BNP levels (n=225) or assessment in a standard manner (n=227). Nonparametric bootstrapping was used to estimate the distribution of incremental costs and effects on the cost-effectiveness plane during 180 days of follow-up.

Results: Testing of BNP induced several important changes in management of dyspnea, including a reduction in the initial hospital admission rate, the use of intensive care, and total days in the hospital at 180 days (median, 10 days [interquartile range, 2-24 days] in the BNP group vs 14 days [interquartile range, 6-27 days] in the control group; $P=.005$). At 180 days, all-cause mortality was 20% in the BNP group and 23% in the control group ($P=.42$). Total treatment cost was significantly reduced in the BNP group ($7930 vs $10 503 in the control group; $P=.004$). Analysis of incremental 180-day cost-effectiveness showed that BNP guidance resulted in lower mortality and lower cost in 80.6%, in higher mortality and lower cost in 19.3%, and in higher or lower mortality and higher cost in less than 0.1% each. Results were robust to changes in most variables but sensitive to changes in rehospitalization with BNP guidance.

Conclusion: Testing of BNP is cost-effective in patients with acute dyspnea.

Arch Intern Med. 2006;166:1081-1087

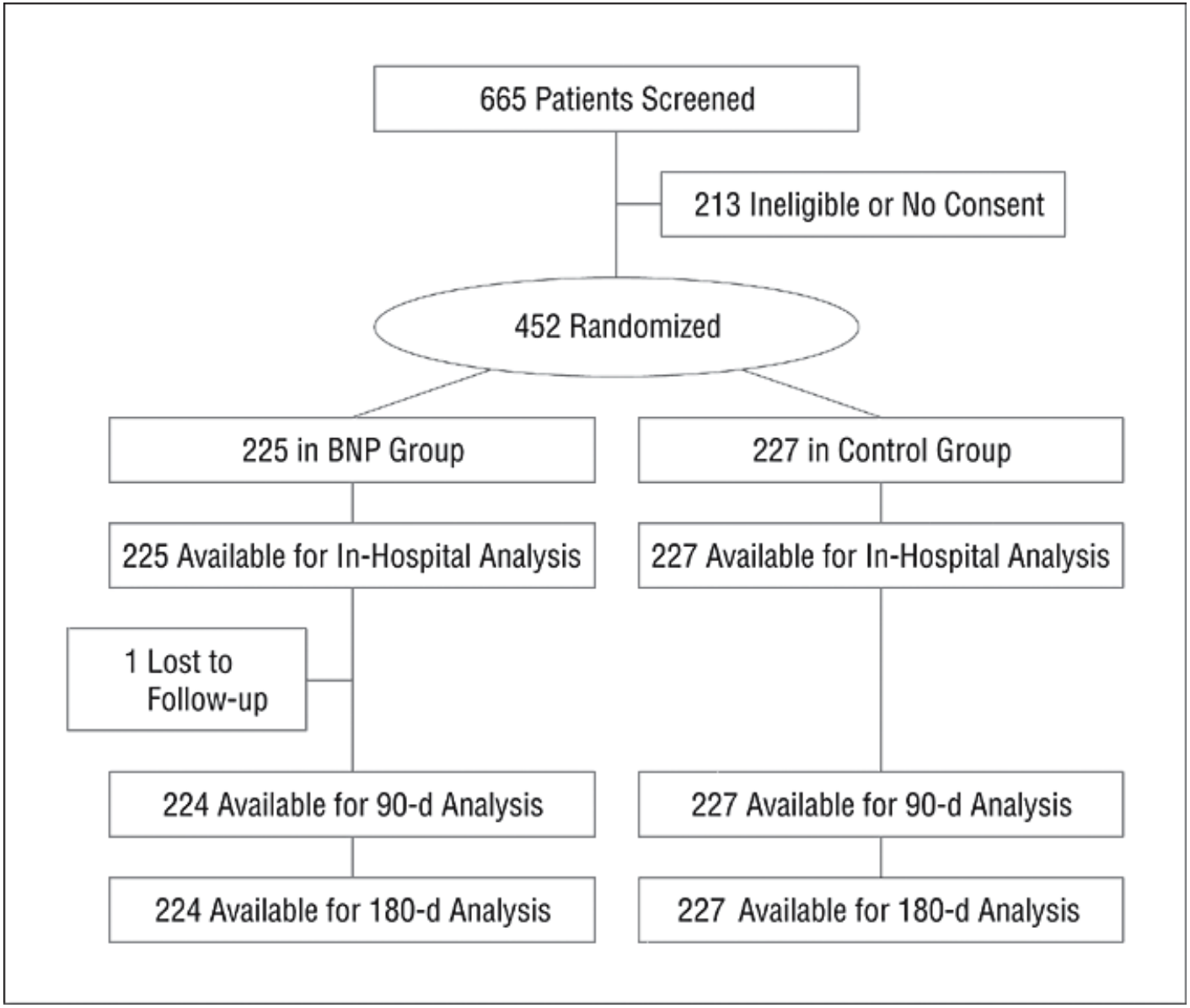
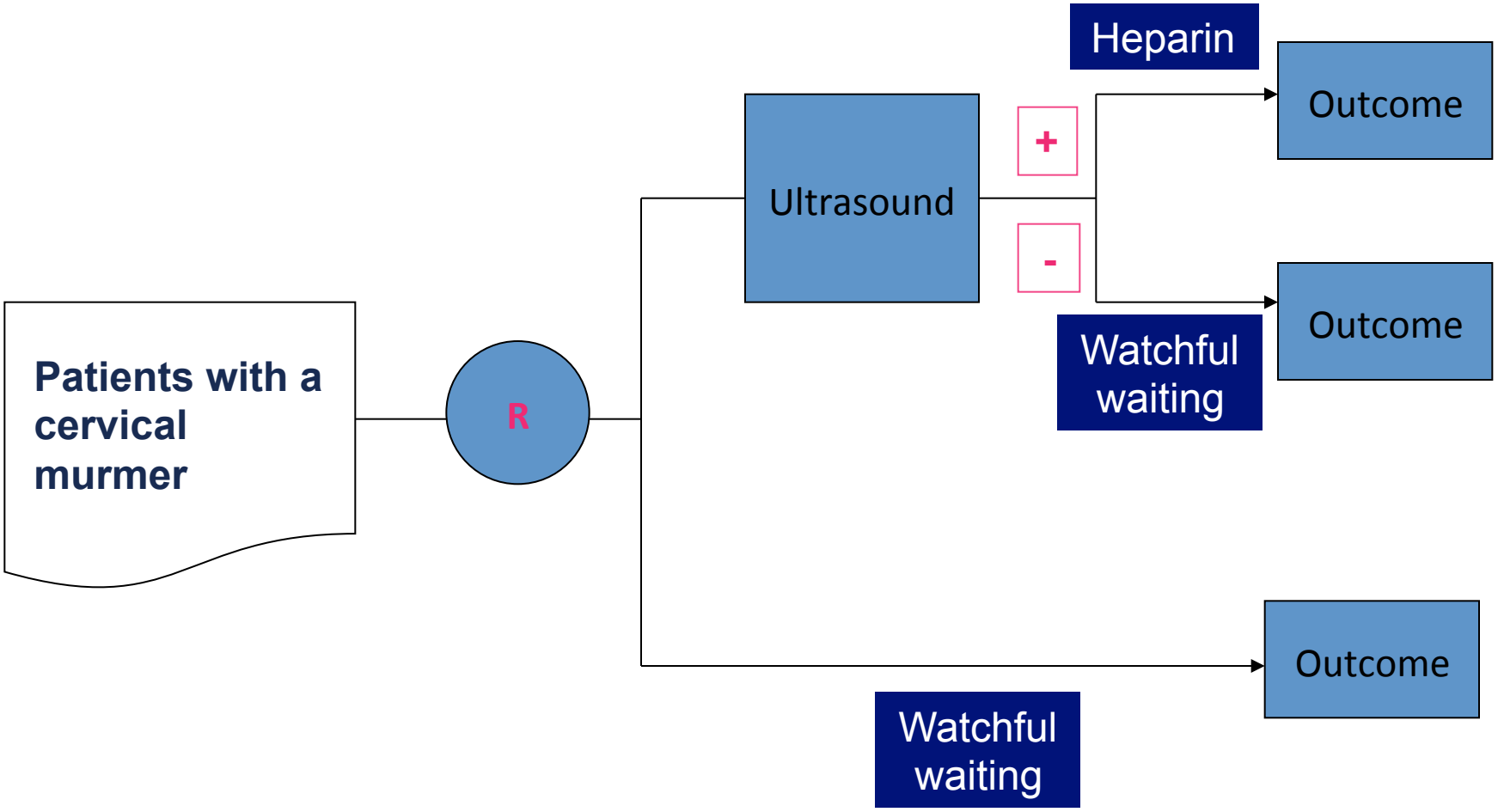

665 Patients Screened
213 Ineligible or No Consent
452 Randomized
225 in BNP Group
227 in Control Group
225 Available for In-Hospital Analysis
227 Available for In-Hospital Analysis
1 Lost to Follow-up
224 Available for 90-d Analysis
227 Available for 90-d Analysis
224 Available for 180-d Analysis
227 Available for 180-d Analysis

Table 2. Outcomes in the BNP and Control Groups

Variable	BNP Group (n = 225)	Control Group (n = 227)	*P* Value
Initial hospital visit			
Total days in hospital, median (IQR)	8 (1-16)	10 (5-18)	.002
If admitted, median (IQR)	11 (6-19)	13 (8-21)	.06
Total treatment cost, mean (SD), $	5410 (6804)	7264 (7363)	.006
All-cause mortality, No. (%)	13 (6)	21 (9)	.21*
At 90 d			
Total days in hospital, median (IQR)	9 (1-19)	13 (6-24)	.001
Days in hospital for dyspnea	8.5 (1-19)	12 (6-23)	.001
Medication cost, mean (SD), $	173 (137)	173 (127)	.98
Total treatment cost, mean (SD), $	6499 (7518)	9037 (8314)	.001
All-cause mortality, No. (%)	32 (14)	36 (16)	.69*
At 180 d			
Total days in hospital, median (IQR)	10 (2-24)	14 (6-27)	.005
Days in hospital for dyspnea	9 (1-20)	13 (6-24)	.003
Medication cost, mean (SD), $	328 (253)	326 (267)	.92
Total treatment cost, mean (SD), $	7930 (8805)	10 503 (10 176)	.004
All-cause mortality, No. (%)	44 (20)	52 (23)	.42*

RCT architecture



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

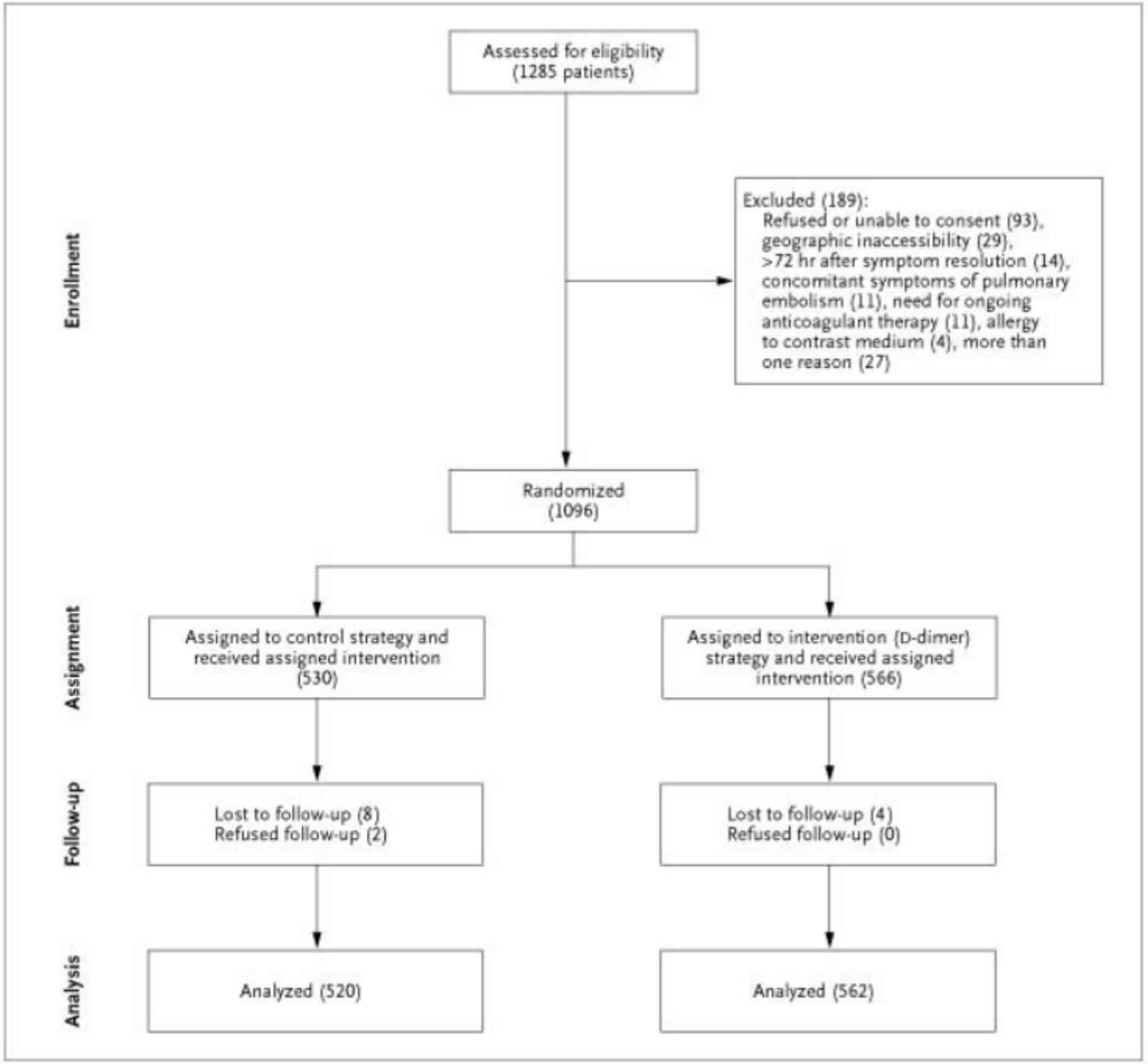
Evaluation of D-Dimer in the Diagnosis of Suspected Deep-Vein Thrombosis

Philip S. Wells, M.D., David R. Anderson, M.D., Marc Rodger, M.D., Melissa Forgie, M.D., Clive Kearon, M.D., Ph.D., Jonathan Dreyer, M.D., George Kovacs, M.D., Michael Mitchell, M.D., Bernard Lewandowski, M.D., and Michael J. Kovacs, M.D.

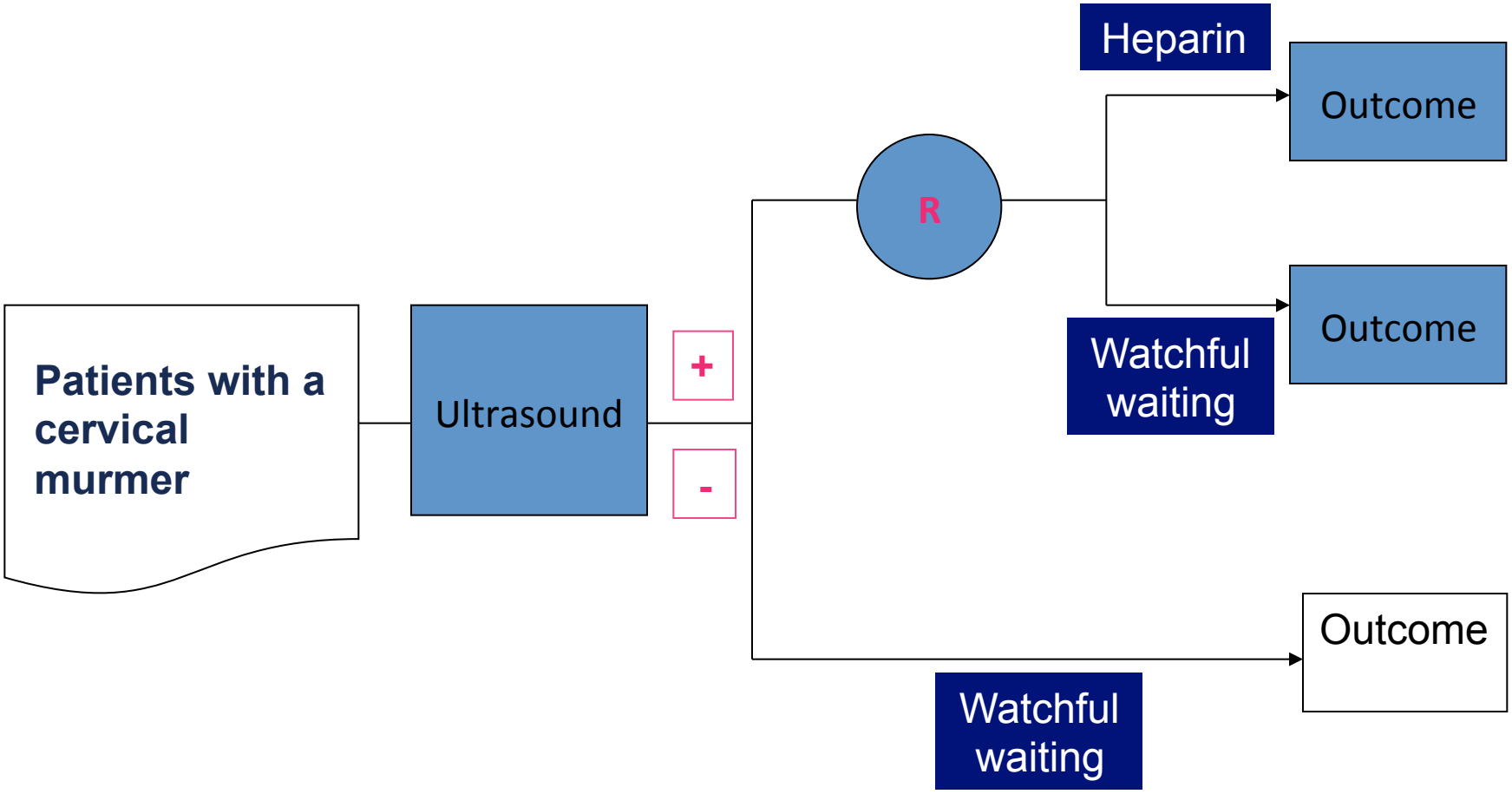
ABSTRACT

BACKGROUND

Several diagnostic strategies using ultrasound imaging, measurement of D-dimer, and assessment of clinical probability of disease have proved safe in patients with suspected deep-vein thrombosis, but they have not been compared in randomized trials.

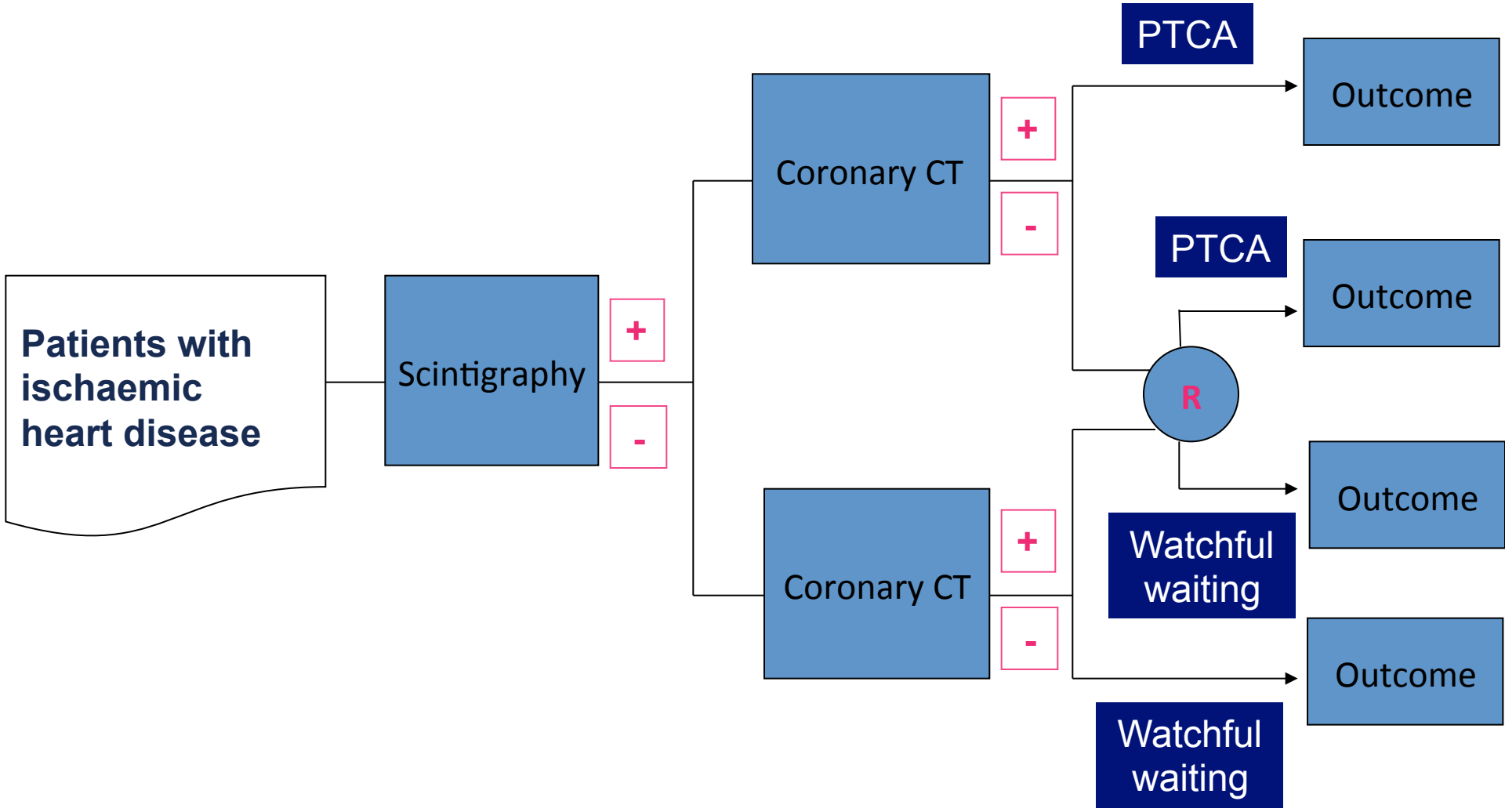

Enrollment
Assessed for eligibility (1285 patients)
Excluded (189):
Refused or unable to consent (93), geographic inaccessibility (29), >72 hr after symptom resolution (14), concomitant symptoms of pulmonary embolism (11), need for ongoing anticoagulant therapy (11), allergy to contrast medium (4), more than one reason (27)
Randomized (1096)
Assignment
Assigned to control strategy and received assigned intervention (530)
Assigned to intervention (D-dimer) strategy and received assigned intervention (566)
Follow-up
Lost to follow-up (8)
Refused follow-up (2)
Lost to follow-up (4)
Refused follow-up (0)
Analysis
Analyzed (520)
Analyzed (562)

RCT architecture



RCT architecture

Let's exercise



Methods

This study was a pragmatic, cluster randomised, factorial, controlled trial. While recognising certain limitations,[39] we chose a cluster randomisation design to optimise the pragmatic nature of the study and to minimise contamination: once general practitioners within a practice had been trained in new communication skills they could not switch at random between using these skills and usual consulting practice. A 2×2 factorial design was used to assess the effect of each intervention and to explore the effect of the interventions combined.[40] Such trials require a prespecified factorial analysis plan with assessments for treatment interactions. We selected this design because we planned to test two treatment hypotheses. The four allocated groups were general practitioners' use of C reactive protein testing (1), training in enhanced communication skills (2), the interventions combined (3), and usual care (4). The groups were combined for analysis as follows: factor A, C reactive protein test (cells 1 and 3) compared with no test (2 and 4) (controlling for the effect of general practitioners' training in enhanced communication skills in the model); and factor B, training in enhanced communication skills (2 and 3) compared with no training (1 and 4) (controlling for the effects of C reactive protein testing in the model).

Outcomes, sample size, and randomisation

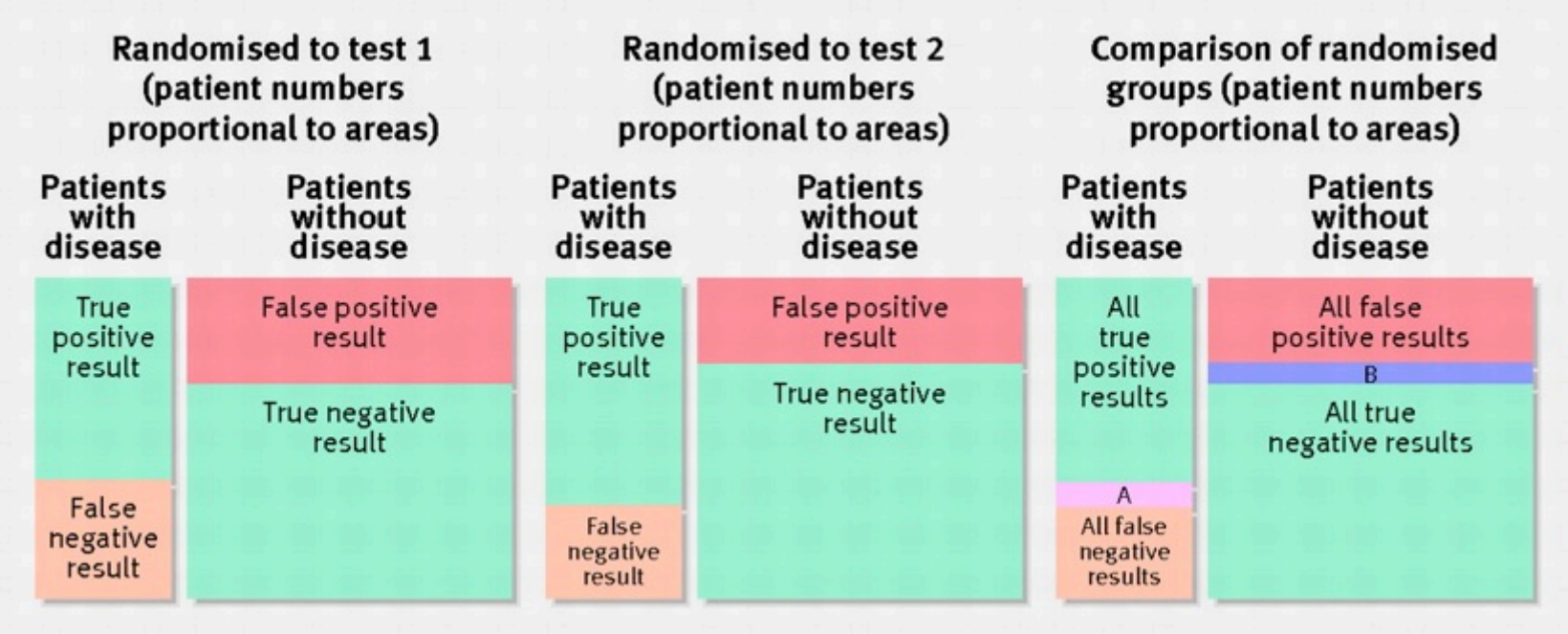
The primary outcome was antibiotic prescribing in the index consultation. Our study required 400 patients with lower respiratory tract infection to detect a reduction in antibiotic prescribing from 80% to 60% (power 80%, α 0.05, follow-up 90%) when adjusted for clustering at practice level (intracluster coefficient 0.06). The sample size was for the main effects only and assumed no interaction between the two interventions. Secondary outcomes were antibiotic prescribing during 28 days' follow-up, reconsultation, clinical recovery, and patients' satisfaction and enablement. Cost effectiveness will be reported separately. We planned to recruit 20 general practices with two participating general practitioners per practice within a large suburban region of the Netherlands. All practices and general practitioners were recruited and provided written consent before randomisation.

Practices were randomised into two groups of 10 practices per intervention, balanced for recruitment potential, resulting in four trial arms (fig 1⇓). The balancing factor used for randomisation was the amount of general practitioners' consultation time (expressed as full time equivalent) that the practice was contributing to the study, and this equated to between one and two full time equivalents for clinical contact time. The randomisation was balanced for those with 1.5 or less full time equivalents and those with more than 1.5 full time equivalents. The Dutch guideline for managing acute cough, including diagnostic and therapeutic advice for lower respiratory tract infection, is distributed to all general practitioners in the Netherlands and informs usual care.[41]

Validity Concerns

- **Blinding**
  - Rare in diagnostic trials (cluster randomisation!)
- **Drop-out**
  - Lack of blinding can induce differential drop-out
  - More stages at which drop-out occurs
- **Compliance**
  - Lack of blinding and complexity in strategies can reduce compliance
- **Power calculations**

Sample size calculations for test-treatment randomised controlled trials.



Ferrante di Ruffano L et al. BMJ 2012;344:bmj.e686

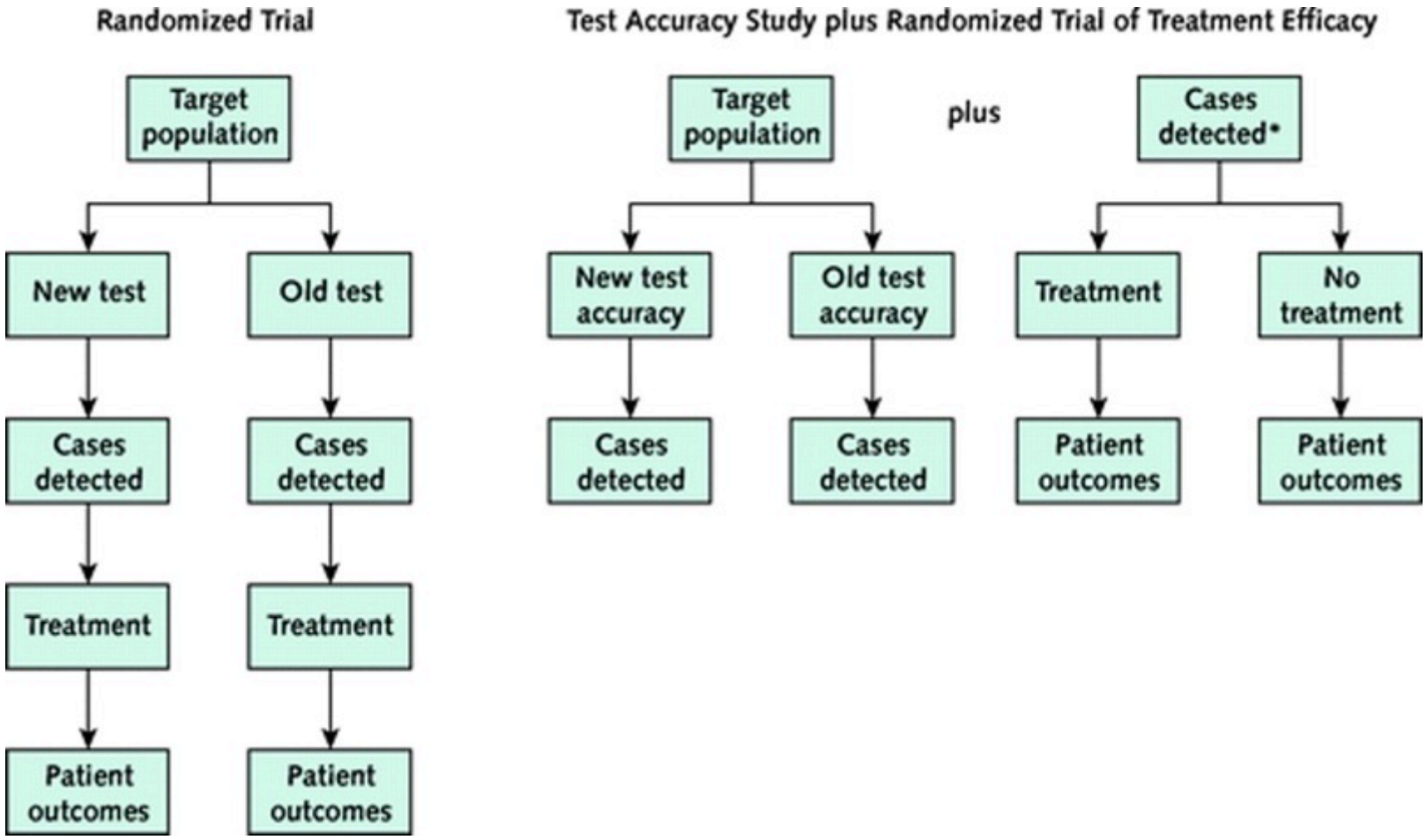
Create your own trial

Modelling

- New test affects patient outcome?
- Only diagnostic accuracy studies
- No trials

- ➔ model impact on patient outcome

Trial evidence versus linked evidence of test accuracy and treatment efficacy



Lord, S. J. et. al. Ann Intern Med 2006;144:850-855

Annals of Internal Medicine

Journal of the American College of Cardiology
Vol. 54, No. 25, 2009
ISSN 0735-1097/09/$36.00
doi:10.1016/j.jacc.2009.10.012

Published by Elsevier Inc.

QUARTERLY FOCUS ISSUE: PREVENTION/OUTCOMES

Clinical Outcomes and Cost-Effectiveness of Coronary Computed Tomography Angiography in the Evaluation of Patients With Chest Pain

Joseph A. Ladapo, MD, PHD,*† Farouc A. Jaffer, MD, PHD,†‡ Udo Hoffmann, MD, MPH,†§ Carey C. Thomson, MD, MPH,†¶ Fabian Bamberg, MD, MPH,§# William Dec, MD,‡ David M. Cutler, PHD,** Milton C. Weinstein, PHD,†† G. Scott Gazelle, MD, MPH, PHD§‖††

Boston and Cambridge, Massachusetts; and Munich, Germany

Objectives The aim of this study was to project clinical outcomes, health care costs, and cost-effectiveness of coronary computed tomography angiography (CCTA), as compared with conventional diagnostic technologies, in the evaluation of patients with stable chest pain and suspected coronary artery disease (CAD).

Background CCTA has recently been found to be effective in the evaluation of patients with suspected CAD, but investigators have raised concerns related to radiation exposure, incidental findings, and nondiagnostic exams.

Methods With published data, we developed a computer simulation model to project clinical outcomes, health care costs, and cost-effectiveness of CCTA, compared with conventional testing modalities, in the diagnosis of CAD. Our target population included 55-year-old patients who present to their primary care physicians with stable chest pain.

Results All diagnostic strategies yielded similar health outcomes, but performing CCTA—with or without stress testing or performing stress single-photon emission computed tomography—marginally minimized adverse events and maximized longevity and quality-adjusted life-years (QALYs). Health outcomes associated with these strategies were comparable, with CCTA in men and women yielding the greatest QALYs but only by modest margins. Overall differences were small, and performing the most effective test—compared with the least effective—decreased adverse event rates by 3% in men and women. Comparable increases in longevity and QALYs were 2 months and 0.1 QALYs in men and 1 month and 0.03 QALYs in women. CCTA raised overall costs, partly through the follow-up of incidental findings, and when performed with stress testing, its incremental cost-effectiveness ratio ranged from $26,200/QALY in men to $35,000/QALY in women. Health outcomes were marginally less favorable in women when radiation risks were considered.

Conclusions CCTA is comparable to other diagnostic studies and might hold good clinical value, but large randomized controlled trials are needed to guide policy. (J Am Coll Cardiol 2009;54:2409–22) © 2009 by the American College of Cardiology Foundation

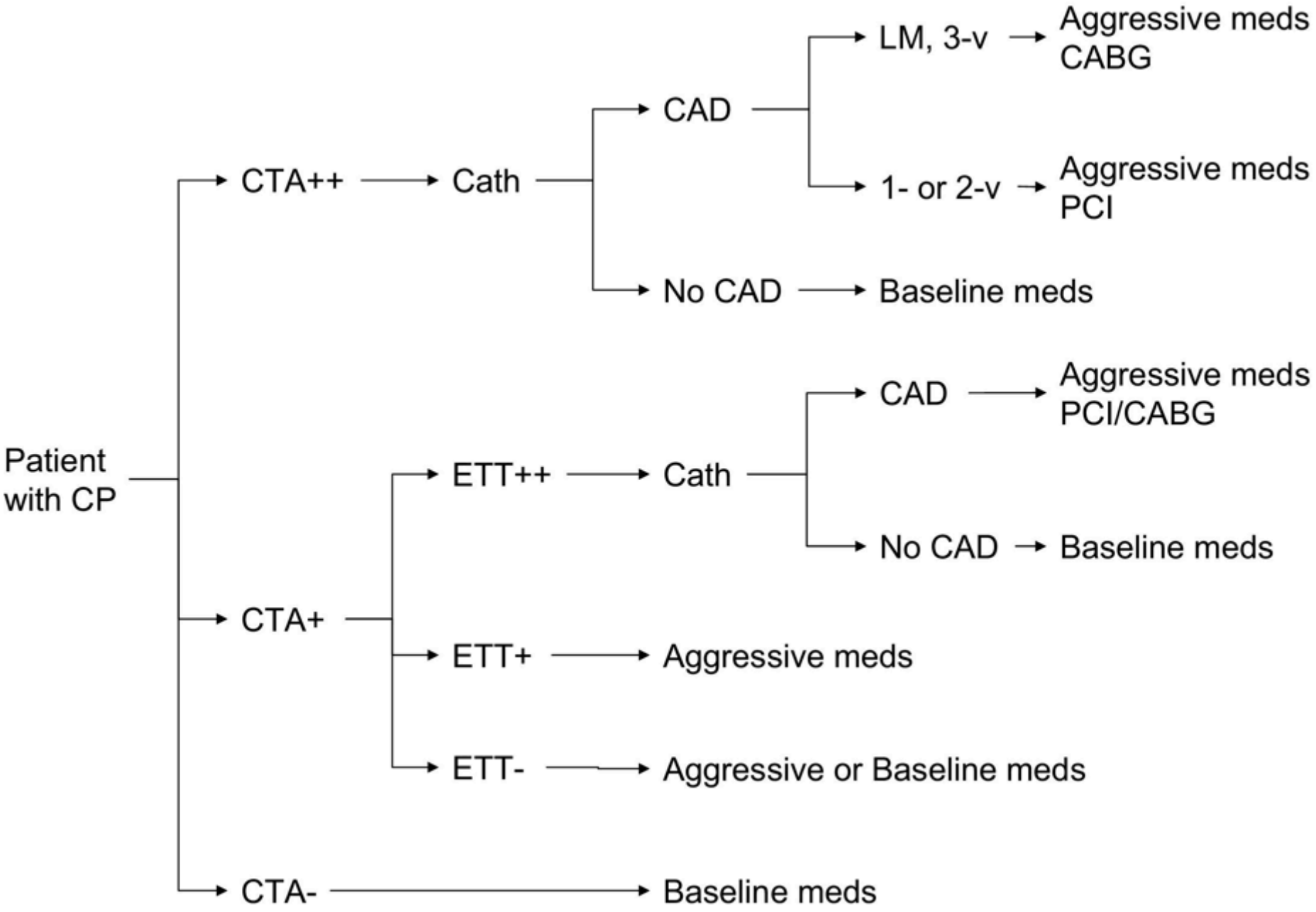

Patient with CP
CTA++
Cath
CAD
LM, 3-v
Aggressive meds CABG
1- or 2-v
Aggressive meds PCI
No CAD
Baseline meds
CTA+
ETT++
Cath
CAD
Aggressive meds PCI/CABG
No CAD
Baseline meds
ETT+
Aggressive meds
ETT-
Aggressive or Baseline meds
CTA-
Baseline meds

Table 4 Clinical Outcomes in 55-Year-Old Men and Women With Chest Pain

	Nonfatal MI*		Nonfatal Stroke*		Life Expectancy, yrs		QALYs	
Test Strategy	Men	Women	Men	Women	Men	Women	Men	Women
CTA–stress ECG	341	192	57	33	77.361	81.633	13.632	16.605
Stress ECG–CTA	350	198	59	34	77.165	81.548	13.552	16.571
CTA	341	192	57	33	77.36	81.633	13.631	16.604
Stress ECG	350	196	59	33	77.198	81.582	13.566	16.582
Stress echocardiography	347	195	59	33	77.247	81.584	13.586	16.585
Stress SPECT	343	193	57	33	77.331	81.628	13.62	16.6
Cardiac catheterization	339	192	57	33	77.316	81.601	13.605	16.588
No exam	380	211	66	37	76.622	81.364	13.33	16.5

*Lifetime prevalence/1,000 patients undergoing diagnostic testing; adverse events only tracked in patients with CAD.
Cath = invasive cardiac catheterization; QALY = quality-adjusted life-year; other abbreviations as in Table 1.

Other patient related outcomes

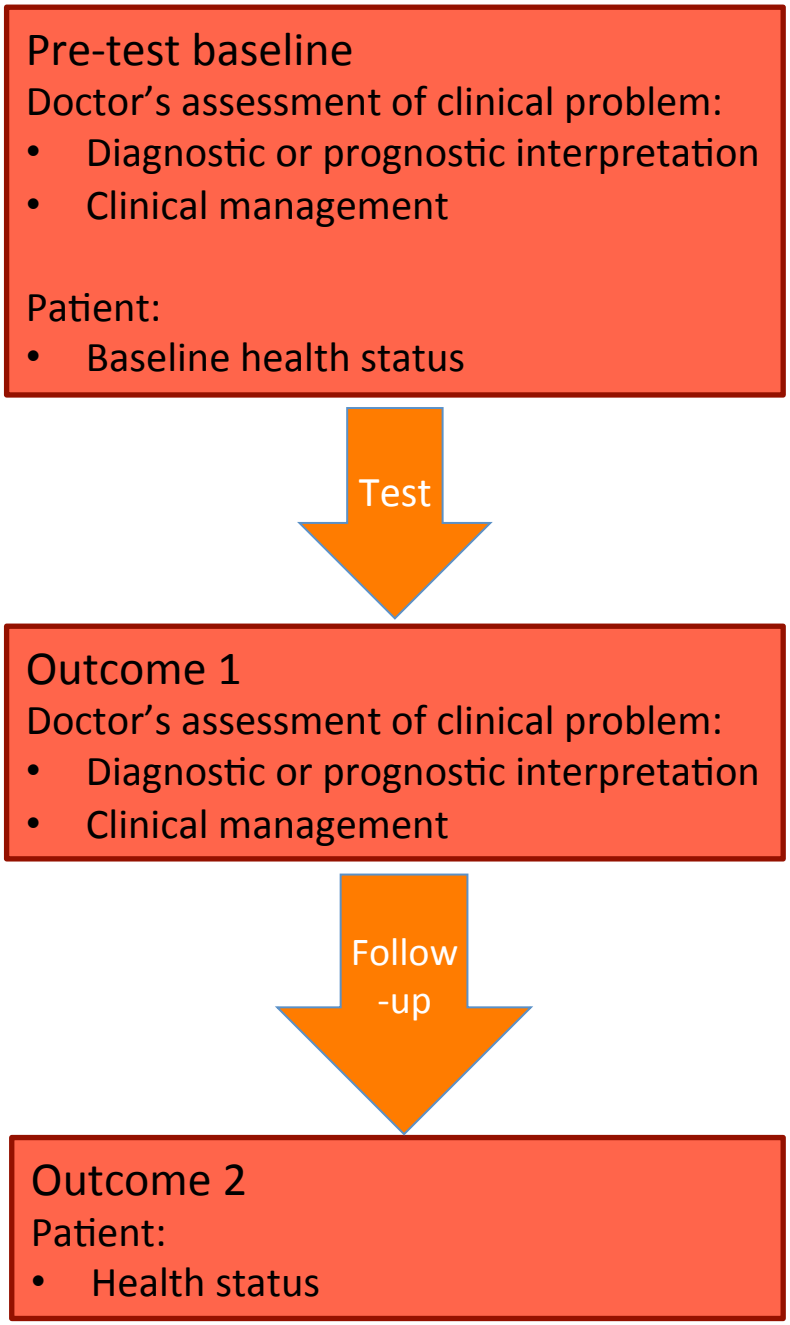
- Pain
- Anxiety
- Guilt

- Insurance problems
- Social isolation
- Placebo effect

=> Qualitative studies, surveys, …

Diagnostic Before-and-After Studies

- To evaluate clinical impact of single or additional testing
- Change in doctor's assessment and management plan
- Impact on clinical course more difficult
  - Long follow-up
  - Interfering factors
- Alternative if RCT impossible, infeasible or unethical


Pre-test baseline
Doctor's assessment of clinical problem:
• Diagnostic or prognostic interpretation
• Clinical management
Patient:
• Baseline health status
Test
Outcome 1
Doctor's assessment of clinical problem:
• Diagnostic or prognostic interpretation
• Clinical management
Follow -up
Outcome 2
Patient:
• Health status

Assessing clinicians' behaviours

- Documentation and standardisation of decision-making
  - Particularly difficult when the comparison group is standard practice
- Assessing behaviour observed in a trial may not be representative
  - Future behaviour will depend on the trial results
  - Learning curves may affect compliance
    - Becoming acquainted with a test
    - Ascertaining how best to use it
    - Gaining confidence in its findings
    - Allowing it to replace other investigations

In summary

- First = what do I want to know about this test?
  - Accuracy?
  - Impact on mortality/morbidity?
  - Impact on other patient related outcomes?
  - Costs?
  - Change in patient pathway?
- Second = design study
- Third = have fun!

Questions?