How to read a paper?

Dr. A M Yusufali, MD, FRCP (Glasg),
Consultant Physician/Cardiologist,
Dubai Hospital, DOHMS

Objective

- Sensitivity and Specificity
- When to use any given diagnostic test?
- Drug trial interpretations
- How to face the drug representative?

How to read a paper: Papers that report diagnostic or screening tests

.

10 men accused of murder

Jury verdict	True criminal status	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

The jury correctly identifies two in every three true murderers;

Jury verdict	True criminal status	
	Murderer	Not murderer
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

It correctly acquits three out of every seven innocent people;

Jury verdict	**True criminal status**	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

If found guilty, there is still only a one in three chance that they are actually a murderer

Jury verdict	True criminal status	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

If found innocent, there is a three in four chance of actually being innocent

Jury verdict	True criminal status	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

In five cases out of every 10 the jury gets it right

Jury verdict	True criminal status	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

Performance of the jury

- Sensitivity: 2/3 = 66.6%
- Specificity: 3/7 = 42.8%
- positive predictive value: 2/6 = 33.3%
- negative predictive value: ¾ = 75%
- Accuracy: 5/10 = 50%

Sensitivity: 2/3 = 66.6% (ability to pick the positive correctly)

Jury verdict	True criminal status	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

Specificity: 3/7 = 42.8%
(ability to pick the negative correctly)

Jury verdict	**True criminal status**	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

positive predictive value: 2/6 = 33.3%
(if +ve the chance that this is correct)

Jury verdict	**True criminal status**	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

negative predictive value: ¾ = 75% (if –ve the chance that that is correct)

Jury verdict	**True criminal status**	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

Accuracy: 5/10 = 50%

Jury verdict	True criminal status	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

Validation study for diagnostic or screening test

	Result of Gold Standard Test	
Result of screening test	Disease positive (A+B)	Disease negative (C+D)
Positive (A+C)	A (True positive)	C (False positive)
Negative (B+D)	B (False negative)	D (True negative)

Validation study of urine glucose testing for diabetes against gold standard

	Result of Gold Standard Test (GTT)	
Result of screening test (Glycosuria)	Disease positive 27	Disease negative 973
Positive 13	6 (True positive)	7 (False positive)
Negative 987	21 (False negative)	966 (True negative)

Calculating the important features of screening test

Feature	Data	Value %
Sensitivity	6/27	22.2
Specificity	966/973	99.3
Positive predictive value	6/13	46.2
Negative predictive value	966/973	97.8
Accuracy	972/1000	97.2

When to use any given test?

- Does prevalence of the disease matter?
- Concept of likelihood ratio
- How to use the pre-test and likelihood ratio to get the post test probability.
- Deciding the cutoff point.
- Another way of looking at the test - ROC

Exploring Accuracy

- accuracy = number of correct diagnoses / number in total population
- a single number that could tell us how well a test performs.
- Let's consider two tests with the same accuracy.
- Let's say we have a population of 1000 patients, of whom 100 have a particular disease (D+).

Two tests with 95.5% accuracy

Test performance: T1 (n=1000)		
	D+	D-
T+	60	5
T-	40	895
PPV = 92.3% NPV = 95.7%		

Test performance: T2 (n=1000)		
	D+	D-
T+	95	60
T-	5	840
PPV = 61.3% NPV = 99.4%		

Prevalence of 1%

Test performance: T1 (n=1000)		
	D+	D-
T+	6	5.5
T-	4	984.5
PPV = 52.2% NPV = 99.6%		

Test performance: T2 (n=1000)		
	D+	D-
T+	9.5	66
T-	0.5	924
PPV = 12.6% NPV = 99.9%		

- If the disease is rare, use of even a very specific test will be associated with many false positives (and all that this entails, especially for a problem like HIV infection);
- conversely, if the disease is common, a positive test is likely to be a true positive.
- You can see from the above that it's rather silly to have a fixed test threshold.

Test with 99% sensitivity and specificity

Test performance: Population A (n=10 000, prevalence 1/1000)		
	D+	D-
T+	**10**	**100**
T-	0	**9890**
PPV = 9.1% NPV = almost 100%		

Test performance: Population B (n=10 000, prevalence 300/1000)		
	D+	D-
T+	2970	70
T-	30	6930
PPV = 97.7% NPV = 99.5%		

Likelihood ratio

- Likelihood that a positive test result would be in a patient with a condition compared to the likelihood that that the same result in a patient without the condition.
- LR = SENSITIVITY / 1- SPECIFICITY

Accuracy = 5/10 = 0.5

Jury verdict	True criminal status	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted	4 Wrongly convicted
Innocent	1 Wrongly acquitted	3 Rightly acquitted

Jury verdict	True criminal status	
	Murderer	**Not murderer**
Guilty	2 Rightly convicted (TP)	4 Wrongly convicted (FP)
Innocent	1 Wrongly acquitted (FN)	3 Rightly acquitted (TN)

Jury verdict	True criminal status	
	Murderer	Not murderer
Guilty	TPF = SENSITIVITY	1-TNF OR FPF
Innocent	1-TPF OR FNF	TNF = SPECIFICITY

Jury guilt test = 0.66/0.57 = 1.16

Jury verdict	True criminal status	
	Murderer	Not murderer
Guilty	Sensitivity = 2/3 = 0.66	1- specificity = 4/7 = 0.57
Innocent	1- sensitivity = 1/3 = 0.33	Specificity = 3/7 = 0.43

Sensitivity = 35/36 =97.2%

RESULT OF CHEST PAIN UNIT	OUTCOME OF GOLD STANDARD TEST	
	POSITIVE	NEGATIVE
POSITIVE	35 Rightly convicted	18 Wrongly convicted
NEGATIVE	1 Wrongly acquitted	238 Rightly acquitted

Specificity = 238/256 = 93%

RESULT OF CHEST PAIN UNIT	OUTCOME OF GOLD STANDARD TEST	
	POSITIVE	NEGATIVE
POSITIVE	35 Rightly convicted	18 Wrongly convicted
NEGATIVE	1 Wrongly acquitted	238 Rightly acquitted

Positive predictive value = 35/53 = 66%

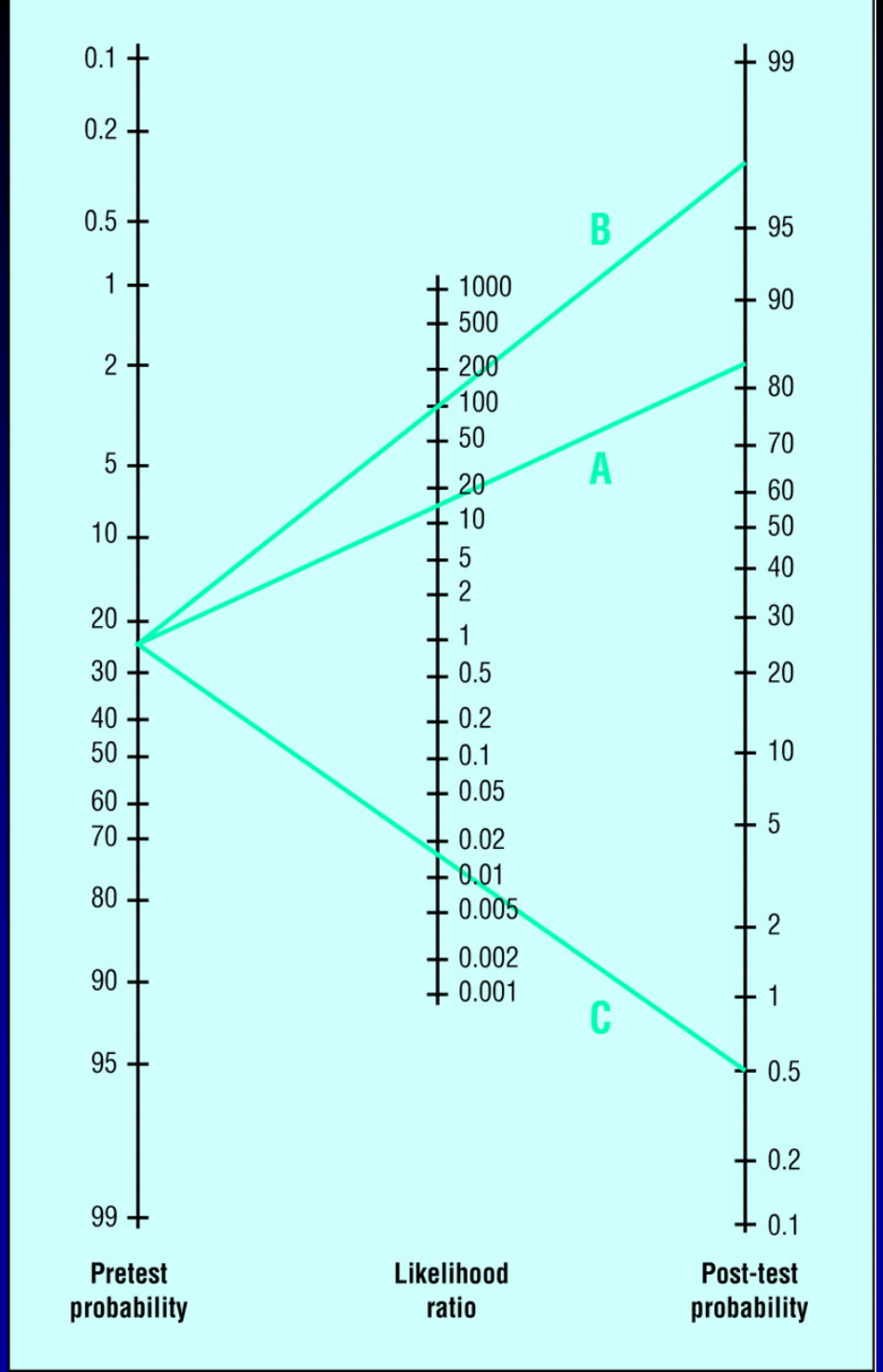
RESULT OF CHEST PAIN UNIT	OUTCOME OF GOLD STANDARD TEST	
	POSITIVE	NEGATIVE
POSITIVE	35 Rightly convicted	18 Wrongly convicted
NEGATIVE	1 Wrongly acquitted	238 Rightly acquitted

Negative predictive value = 238/239 =99.6%

RESULT OF CHEST PAIN UNIT	OUTCOME OF GOLD STANDARD TEST	
	POSITIVE	NEGATIVE
POSITIVE	35 Rightly convicted	18 Wrongly convicted
NEGATIVE	1 Wrongly acquitted	238 Rightly acquitted

Likelihood ratio = 0.972/0.07 = 13.8

RESULT OF CHEST PAIN UNIT	OUTCOME OF GOLD STANDARD TEST	
	POSITIVE	**NEGATIVE**
POSITIVE	Sensitivity = 0.972	1-specificity = 0.07
NEGATIVE	1- sensitivity = 0.028	Specificity = 0.93


0.1
0.2
0.5
1
2
5
10
20
30
40
50
60
70
80
90
95
99
1000
500
200
100
50
20
10
5
2
1
0.5
0.2
0.1
0.05
0.02
0.01
0.005
0.002
0.001
99
95
90
80
70
60
50
40
30
20
10
5
2
1
0.5
0.2
0.1
A
B
C
Pretest probability
Likelihood ratio
Post-test probability

Another example using likelihood ratios

- prostate specific antigen (PSA) test to screen for prostate cancer.
- Most men will have some detectable antigen in their blood (say, 0.5 ng/ml), and most of those with advanced prostate cancer will have high concentrations (above about 20 ng/ml).
- But a concentration of, say, 7.4 ng/ml may be found either in a perfectly normal man or in someone with early cancer.
- There simply is not a clean cutoff between normal and abnormal

- We can, however, use the results of a validation study of this test against a gold standard for prostate cancer (say a biopsy of the prostate gland) to draw up a whole series of two by two tables.
- Each table would use a different definition of an abnormal test result to classify patients as "normal" or "abnormal."
- From these tables, we could generate different likelihood ratios associated with an antigen concentration above each different cutoff point.
- When faced with a test result in the "grey zone" we would at least be able to say, "This test has not proved that the patient has prostate cancer, but it has increased [or decreased] the odds of that diagnosis by a factor of *x*."

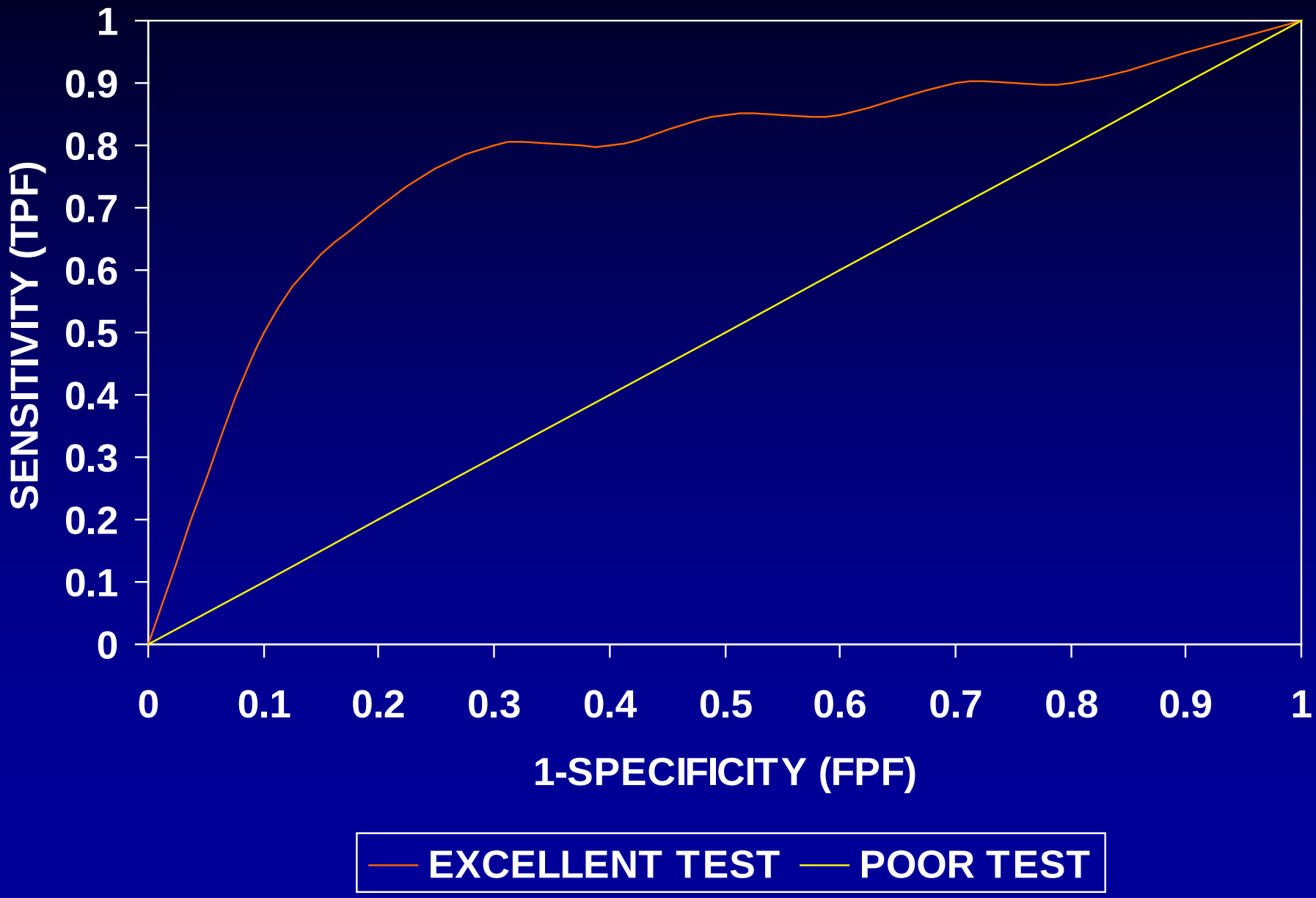
Where should the cutoff point be?

- Let’s consider the diagnosis of hypothyroidism
- Test used is TSH. Why?
- Why 5 is taken as a cutoff point?
- Why not 2,3,4,5,6,7,8
- If we take 8 as cutoff point, the all >8 will definitely be hypothyroid. (high specificity)
- However we will miss a lot (false negative)
- If 3 is cutoff, a lot of false positive (very sensitive)

Deciding on a test threshold

- Financial costs, both direct and indirect of treating a disease (present or not), and of failing to treat a disease;
- Costs of further investigation (where deemed appropriate);
- Discomfort to the patient caused by disease treatment, or failure to treat;
- Mortality associated with treatment or non-treatment;

RECIEVER OPERATOR CHARACTERISTIC CURVE



Where should the cutoff point be?

- If the cost of missing a diagnosis is great, and treatment (even inappropriate treatment of a normal person) is safe, then one should move to a point on the *right* of the ROC, where we have a high TPF (most of the true positives will be treated) at the cost of many false positives.
- Conversely, if the risks of therapy are grave, and therapy doesn't help much anyway, we should position our point far to the left, where we'll miss a substantial number of positives (low TPF) but not harm many unaffected people (low FPF)!

Summary points

- New tests should be validated by comparison against an established gold standard in an appropriate spectrum of subjects
- Diagnostic tests are seldom 100% accurate (false positives and false negatives will occur)
- A test is valid if it detects most people with the target disorder (high sensitivity) and excludes most people without the disorder (high specificity), and if a positive test usually indicates that the disorder is present (high positive predictive value) and a high likelihood ratio.
- ROC area should be nearer to 1

How to read a paper: Papers that report drug trials

Objective

- The bottom line
- Survival analysis
- Hazard ratio

The bottom line

- Is "statistically significant difference" enough?
- Four simple calculations will enable you to answer this question objectively and in a way that means something to the non-statistician.

These calculations are:

- the relative risk reduction,
- the absolute risk reduction,
- the number needed to treat, and
- the odds ratio.

coronary artery surgery triallists collaboration.
Lancet 1994;344:563-70

treatment and outcome		
	Outcome at 10 years	
Treatment	**Dead**	**Alive**
Medical treatment (n=1325)	404	921
Coronary artery bypass grafting (n=1324)	350	974

Patients receiving medical treatment have a chance of 404/1325=0.305 or 30.5% of being dead at 10 years. Let us call this risk *x*.

Patients randomized to coronary artery bypass grafting have a chance of 350/1324=0.264 or 26.4% of being dead at 10 years. Let us call this risk *y*.

treatment and outcome		
	Outcome at 10 years	
Treatment	**Dead**	**Alive**
Medical treatment (n=1325)	404	921
Coronary artery bypass grafting (n=1324)	350	974

The relative risk of death :—that is, the risk in surgically treated patients compared with medically treated controls—is *y/x* or 0.264/0.305=0.87 (87%).

treatment and outcome		
	Outcome at 10 years	
Treatment	**Dead**	**Alive**
Medical treatment (n=1325)	404	921
Coronary artery bypass grafting (n=1324)	350	974

The relative risk reduction: —that is, the amount by which the risk of death is reduced by the surgery—is 100%-87% (1-*y*/*x*)=13%.

The absolute risk reduction (or risk difference): —that is, the absolute amount by which surgical treatment reduces the risk of death at 10 years—is 30.5%-26.4%=4.1% (0.041).

treatment and outcome		
	Outcome at 10 years	
Treatment	**Dead**	**Alive**
Medical treatment (n=1325)	404	921
Coronary artery bypass grafting (n=1324)	350	974

The number needed to treat: —how many patients need coronary artery bypass grafting in order to prevent, on average, one death after 10 years—is the reciprocal of the absolute risk reduction: 1/ARR=1/0.041=24.

Odds ratio: the "odds" of dying compared with the odds of surviving for patients in the medical treatment group is 404/921=0.44, and for patients in the surgical group is 350/974=0.36. The ratio of these odds will be 0.36/0.44=0.82.

Calculating the "bottom line" effects on an intervention

	Outcome event		
Group	**Yes**	**No**	**Total**
Control group	a	b	a+b
Experimental group	c	d	c+d

Control event rate (CER)=risk of outcome event in control group=a/(a+b)

Experimental event rate (EER)=risk of outcome event in experimental group=c/(c+d)

Relative risk reduction (RRR)=(CER—EER)/CER

Absolute risk reduction (ARR)=CER—EER

Number needed to treat (NNT)=1/ARR=1/(CER—EER)

$$\text{Odds ratio} = \frac{\text{(odds of outcome event } v \text{ odds of no event) in intervention group}}{\text{(odds of outcome event } v \text{ odds of no event) in control group}}$$

Survival analysis

- Time until a single event occurs
- Most time death but could be discharge from hospital, stroke, myocardial infarc, fracture etc.
- Able to deal with situations in which the end event has not happened in every patient or when information on a case is only known for a limited duration – known as ‘censored’ observations.

Life table

- Is a table of the proportion of the patients surviving over time.
- Life table methods look at the data at a number of fixed time points and culculate the survival rate at those times.
- The most commonly used method is Kaplan-Meier.

Kaplan-Meier

- The Kaplan-Meier approach recalculates the survival rate when the end event occurs in the data set, when a change happens rather than at fixed intervals.
- This is usually represented as a “survival plot”.
- The plots comparing two or more groups make any differences between the groups beutifully clear.
- The test to compare the two groups is called the “log rank test”.
- Its *P* value will tell you how significant the result of the test is.

TABLE 3. INCIDENCE OF THE PRIMARY OUTCOME AND OF DEATHS FROM ANY CAUSE.

OUTCOME	RAMIPRIL GROUP (N=4645)	PLACEBO GROUP (N=4652)	RELATIVE RISK (95% CI)*	Z STATISTIC	P VALUE†
	no. (%)				
Myocardial infarction, stroke, or death from cardiovascular causes‡	651 (14.0)	826 (17.8)	0.78 (0.70–0.86)	−4.87	<0.001
Death from cardiovascular causes§	282 (6.1)	377 (8.1)	0.74 (0.64–0.87)	−3.78	<0.001
Myocardial infarction§	459 (9.9)	570 (12.3)	0.80 (0.70–0.90)	−3.63	<0.001
Stroke§	156 (3.4)	226 (4.9)	0.68 (0.56–0.84)	−3.69	<0.001
Death from noncardiovascular causes	200 (4.3)	192 (4.1)	1.03 (0.85–1.26)	0.33	0.74
Death from any cause	482 (10.4)	569 (12.2)	0.84 (0.75–0.95)	−2.79	0.005

*CI denotes confidence interval.

†P values were calculated with use of the log-rank test.

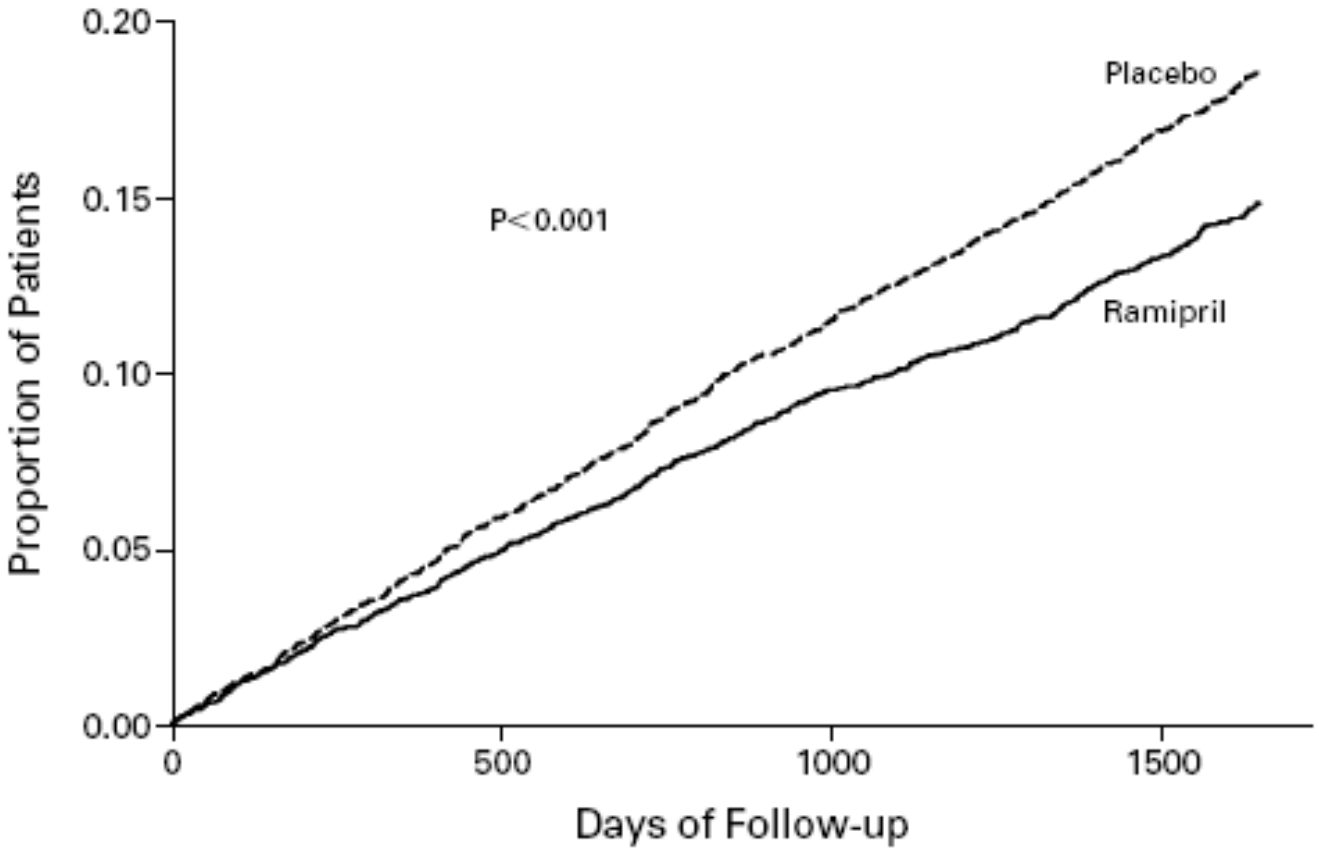


Figure 1. Kaplan–Meier Estimates of the Composite Outcome of Myocardial Infarction, Stroke, or Death from Cardiovascular Causes in the Ramipril Group and the Placebo Group.

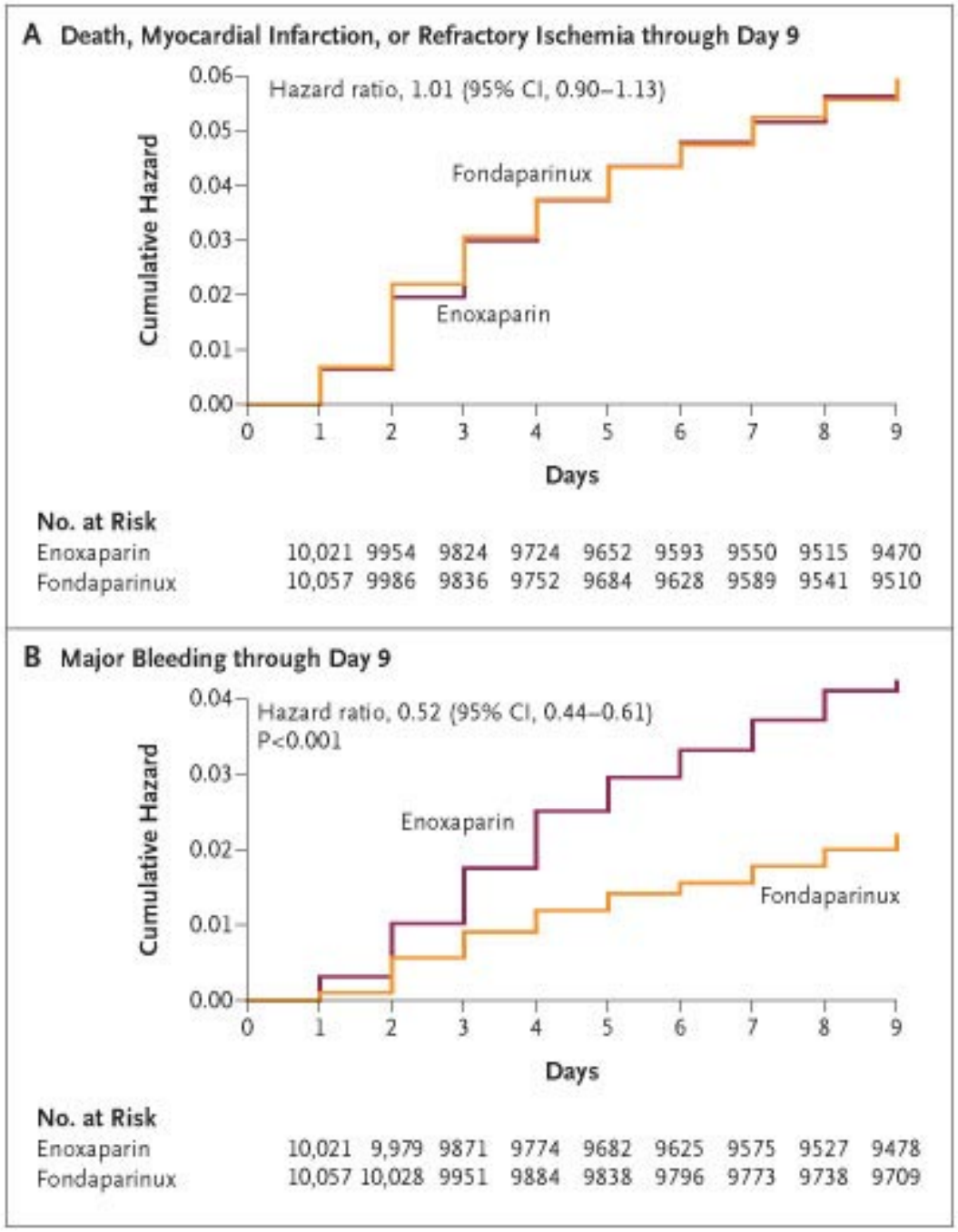
The relative risk of the composite outcome in the ramipril group as compared with the placebo group was 0.78 (95 percent confidence interval, 0.70 to 0.86).

Hazard ratio

- The COX REGRESSION MODEL is used to investigate the relationship between an event (usually death) and possible explanatory variables, for instance smoking status or weight.
- It provides us with estimates of the effect that different factors have on the time until the end event.

Hazard ratio

- The ‘HR’ is the ratio of the hazard (chance of something harmful happening) of an event in one group of observations divided by the hazard of an event in another group.
- A HR of 1 means the risk is 1x that of the second group, i.e. the same.
- A HR of 2 implies twice the risk.


A Death, Myocardial Infarction, or Refractory Ischemia through Day 9
Hazard ratio, 1.01 (95% CI, 0.90–1.13)
Fondaparinux
Enoxaparin
Cumulative Hazard
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0 1 2 3 4 5 6 7 8 9
Days
No. at Risk
Enoxaparin 10,021 9954 9824 9724 9652 9593 9550 9515 9470
Fondaparinux 10,057 9986 9836 9752 9684 9628 9589 9541 9510
B Major Bleeding through Day 9
Hazard ratio, 0.52 (95% CI, 0.44–0.61)
P<0.001
Enoxaparin
Fondaparinux
Cumulative Hazard
0.04
0.03
0.02
0.01
0.00
0 1 2 3 4 5 6 7 8 9
Days
No. at Risk
Enoxaparin 10,021 9,979 9871 9774 9682 9625 9575 9527 9478
Fondaparinux 10,057 10,028 9951 9884 9838 9796 9773 9738 9709

How to face a drug representative?

- Drug representative
- What areas we should be interested in ? STEP
- Surrogate end points
- Statistics at work

“drug rep”

- See representatives only by appointment. Choose to see only those whose product interests you, and confine the interview to that product
- Take charge of the interview. Do not hear out a rehearsed sales routine but ask directly for the information below
- Request independent published evidence from reputable, peer reviewed journals
- Do not look at promotional brochures, which may contain unpublished material, misleading graphs, and selective quotations
- Ignore anecdotal "evidence," such as the fact that a medical celebrity is prescribing the product
- Using the STEP acronym, ask for evidence in four specific areas:

"Evidence" and marketing

- the pharmaceutical industry is interested in you and is investing a staggering sum of money trying to influence you.
- a briefcase full of "evidence" in support of their wares.
- Pharmaceutical "reps" are now much more informative than they used to be, but they may show ignorance of basic epidemiology and clinical trial design

STEP

- *Safety*—the likelihood of long term or serious side effects caused by the drug (remember that rare but serious adverse reactions to new drugs may be poorly documented)
- *Tolerability*—best measured by comparing the pooled withdrawal rates between the drug and its most significant competitor
- *Efficacy*—the most relevant dimension is how the product compares with your current favourite
- *Price*—should take into account indirect as well as direct costs

How to get evidence out of a drug rep

- Evaluate the evidence stringently, paying particular attention to the power (sample size) and methodological quality of clinical trials, and the use of surrogate end points. Do not accept theoretical arguments in the drug's favour ("longer half life," for example) without direct evidence that this translates into clinical benefit
- Do not accept the newness of a product as an argument for changing to it. Indeed, there are good scientific arguments for doing the opposite
- Decline to try the product via starter packs or by participating in small scale, uncontrolled "research" studies
- Record in writing the content of the interview and return to these notes if the "rep" requests another audience

Surrogate end points

- can considerably reduce the sample size, duration, and, therefore, cost, of clinical trials;
- can allow treatments to be assessed in situations where the use of primary outcomes would be excessively invasive or unethical

But surrogate end points have some drawbacks.

- Firstly, does not itself answer the essential preliminary questions: "what is the objective of treatment in this patient?"
- Secondly, it may not be valid or reliable.
- Thirdly, over reliance on a single surrogate end point narrow clinical perspective.

Use of ramipril in preventing stroke: double blind randomised trial

Jackie Bosch, Salim Yusuf, Janice Pogue, Peter Sleight, Eva Lonn, Badrudin Rangoonwala, Richard Davies, Jan Ostergren, Jeff Probstfield on behalf of the HOPE Investigators

Abstract

Objective: To determine the effect of the angiotensin converting enzyme inhibitor ramipril on the secondary prevention of stroke.

Design: Randomised controlled trial with 2 x 2 factorial design.

Setting: 267 hospitals in 19 countries.

Participants: 9297 patients with vascular disease or diabetes plus an additional risk factor, followed for 4.5 years as part of the HOPE study.

Outcome measures

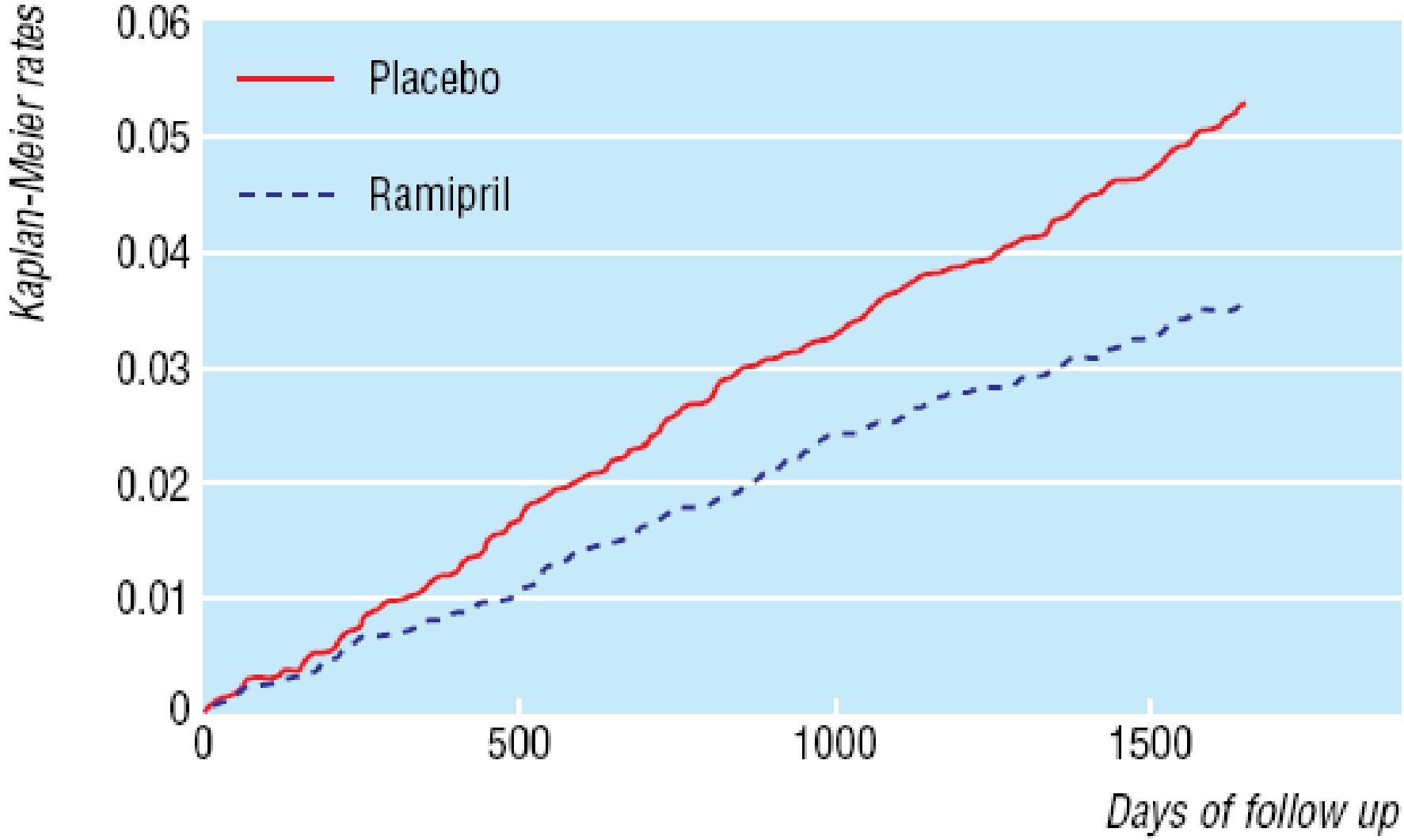
Stroke (confirmed by computed tomography or magnetic resonance imaging when available), transient ischaemic attack, and cognitive function. Blood pressure was recorded at entry to the study, after 2 years, and at the end of the study.

Results

- Reduction in blood pressure was modest (3.8mm Hg systolic and 2.8 mm Hg diastolic).
- The relative risk of any stroke was reduced by 32% (156/4645 *v* 226/4652) in the ramipril group compared with the placebo group, and the relative risk of fatal stroke was reduced by 61% (17/4645 *v* 44/4652).
- Benefits were consistent across baseline blood pressures, drugs used, and subgroups defined by the presence or absence of previous stroke, coronary artery disease, peripheral arterial disease, diabetes, or hypertension.
- Significantly fewer patients on ramipril had cognitive or functional impairment.

Conclusion

Ramipril reduces the incidence of stroke in patients at high risk, despite a modest reduction in blood pressure.



Kaplan-Meier estimates of the development of stroke by treatment group. The relative risk of developing stroke in the ramipril group compared with the placebo group was 0.68 (95% confidence interval 0.56 to 0.84; P=0.0002).

What statistical method was used and why?

- *Null hypothesis*: was that there was no difference between the groups
- The *Kaplan-Meier plot* was used to give a visual representation of the time until having a stroke in each group.
- The survival between the two groups was compared using the *log rank test*.

What do the results mean?

These calculations are:

- the relative risk reduction,
- the absolute risk reduction,
- the number needed to treat, and
- the odds ratio.

Results

treatment and outcome		
Treatment	**Outcome at 4.5 years**	
	stroke	No stroke
Ramipril (4645)	156	4489
Placebo (4652)	226	4426

Patients receiving Ramipril have a risk of getting stroke = 156/4645 = 3.36%
Patients receiving Placebo have a risk of getting stroke = 226/4652 = 4.86%
The Absolute Risk Reduction (ARR) = 4.86-3.36 =1.5%
The Relative Risk Reduction (RRR) = 1.5/4.86 =32%
The Number Needed to Treat (NNT) = 100/ARR = 100/1.5 = 67

treatment and outcome		
Treatment	**Outcome at 4.5 years**	
	stroke	No stroke
Ramipril (4645)	156	4489
Placebo (4652)	226	4426