Should we screen for breast cancer and what do we tell women about the effectiveness of screening?

Epidemiology and Statistics Module
Lecture 5

Sandra Eldridge

Aims of the lecture

- Recap on incidence and prevalence
- Discuss data sources including cancer registries
- Explain what is meant by sensitivity, specificity, and positive and negative predictive value and how these relate to the prevalence of the condition
- Explain what is meant by lead time, lead time bias and overdiagnosis
- Describe controversies over screening
- Describe the Wilson and Jungner criteria

Using data and quotes from the Wainer paper….

How common is breast cancer?

Lifetime prevalence:

"one in every eight US women can expect to be diagnosed with breast cancer"

Number of cancers (incidence):

"180,000 new cases per year of invasive breast cancer are diagnosed in women in the United States"

Annual incidence rate, USA:

(http://www.wcrf-uk.org/research/cancer_statistics)

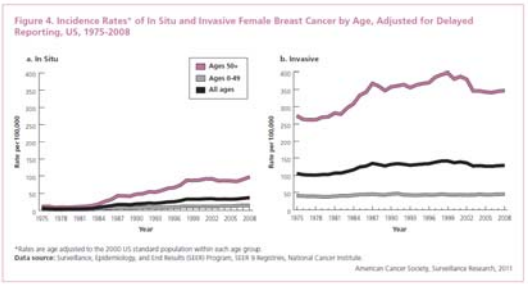
76 per 100,000

Incidence varies by age group

Sources of data on cancer incidence and mortality

- Cancer registries cover all of US population
- 11 cancer registries in the UK
  - Collect data from a variety of sources
  - Aim: to deliver timely, comparable and high-quality cancer data
  - Advantages / disadvantages?
- Registries in other countries?
- Studies undertaken by researchers

Trends in breast cancer in the USA



Should we screen?

- How good is the screening test?
- Screening test is not 100% accurate

  "We are told that accuracy of mammograms varies from 80% to 90% depending on circumstances"
- "Accuracy" = how good the test is at identifying those with or without cancer
- Technical terms - sensitivity & specificity

Sensitivity and specificity

Test says they do = 90%

Sensitivity

People who really **do** have breast cancer

Test says they don't = 90%

Specificity

People who really **don't** have breast cancer

Sensitivity and specificity – what they don't tell us

- If you have a positive test how likely is it that you have cancer?
- If you have a negative test how likely is it that you don't have cancer?
- The answers to these questions depend on:
  - Sensitivity
  - Specificity
  - Prevalence

Two by two table

Sensitivity = 90%
=180,000/?
=180,000/200,000

Assumes all those diagnosed are as a result of screening

	Breast cancer	Not breast cancer	TOTALS
Positive test	180,000		
Negative test			
TOTALS			33,500,000

Specificity = 90%
=?/33,300,000
= 29,970,00/33,300,000

Two by two table

	Breast cancer	Not breast cancer	TOTALS
Positive test	180,000	3,330,000	3,510,000
Negative test	20,000	29,970,000	29,990,000
TOTALS	200,000	33,300,000	33,500,000

Two by two table

Likelihood of cancer if positive test = positive predictive value
=180000/3510000=5%

	Breast cancer	Not breast cancer	TOTALS
Positive test	180,000	3,330,000	3,510,000
Negative test	20,000	29,970,000	29,990,000
TOTALS	200,000	33,300,000	33,500,000

Likelihood of not having cancer if negative test = negative predictive value =29970000/29990000=99.9%

Positive and negative predictive value

If I have a positive test what is the likelihood that I have breast cancer? Positive predictive value = 1/20 = 5%

If I have a negative test what is the likelihood that I don't have breast cancer? Negative predictive value is over 99/100 = 99%

Sensitivity, specificity, positive and negative predictive values can also be used for diagnosis

Compare with diagnosis for whooping cough from last week

Barts and The London School of Medicine and Dentistry www.smd.qmul.ac.uk

How did Harnden et al. confirm that the children had whooping cough?

- Serology (blood test)
  - 100 ELISA units/ml IgG concentration used as cut-off point for a positive diagnosis
- How good is serology for diagnosis?
- Are children ever misdiagnosed?

Barts and The London School of Medicine and Dentistry www.smd.qmul.ac.uk

Diagnosing whooping cough amongst those with persistent cough

- Laboratory tests
  - Culture
  - PCR
  - Serology
- Symptoms
  - Paroxysmal cough
  - Vomiting
  - Whooping

(for more detail see Harnden, 2010)

Diagnosis using serology

Test says they do = 76%

Sensitivity

People who really **do** have whooping cough

Test says they don't = 99%

Specificity

People who really **don't** have whooping cough

Whooping cough diagnosis using serology

	Whooping cough	Not whooping cough	TOTALS
Positive test	370		
Negative test			
TOTALS			1000

PPV=370/375 = 99%

NPV = 508/625 = 81%

Whooping cough diagnosis using serology

	Whooping cough	Not whooping cough	TOTALS
Positive test	370	5	375
Negative test	117	508	625
TOTALS	487	513	1000

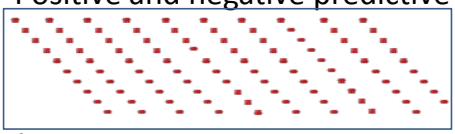
PPV=370/375 = 99%

NPV = 508/625 = 81%

Whooping cough diagnosis

- Focused population (school children aged 5 to 16 with persistent cough)
- More likely to have whooping cough
- Therefore greater prevalence
- Therefore higher PPV and lower NPV

Positive and negative predictive value



If I have a positive test what is the likelihood that I have whooping cough? Positive predictive value = over 99%

If I have a negative test what is the likelihood that I don't have whooping cough? Negative predictive value =81%

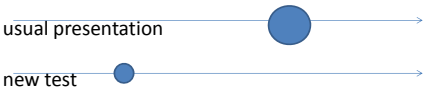
Breast cancer - what happens after screening?

- True positives 180,000 – treated, possible better prognosis?
- False positives 3,330,000 – identified as such?, treated?
- True negatives 29,970,000 – nothing
- False negatives 20,000 – falsely reassured, diagnosed later?

IF LOTS OF FALSE POSITIVES, THEIR OUTCOMES ARE IMPORTANT

Should we screen?

- Screening detects disease earlier
- How do you measure increase in life expectancy?



- **Lead time** is the length of time between the detection of a disease (usually based on new, experimental criteria) and its usual clinical presentation and diagnosis (based on traditional criteria)
- **Lead time bias** is the bias arising in survival analyses when the disease is detected earlier in one group than the other

Life expectancy and lead time bias

- Expected survival from detection
  - 5 years if tumour detected through self-examination
  - 10 years if tumour detected when earlier through screening
- Conclusion = detection increases survival by 5 years??
- How long does it take for tumour to get from being detectable through screening to detectable through self examination?

Should we screen?

"Furthermore, a majority of tiny tumours never grow to become life-threatening. Instead they just sit there and do nothing."

- Natural history
- Screening detects something which may disappear or may never progress to cause problems in a patient's lifetime **(overdiagnosis)**
- Should we subject women to unpleasant treatments?

Should we screen?

It costs lots of money (see arguments in Wainer paper)

Should we screen?

- We need to take account of trends, other factors *"Breast cancer incidence and mortality rates remain highest in developed countries compared with developing countries, as a result of differential use of screening mammograms and disparities in lifestyle and hereditary factors"* (Althius 2005)
- What's the evidence?
- How was it obtained?
- What does it mean about screening?

Should we screen?

- Overall, do those who are screened have better outcomes than those who do not?
- How do we answer this?
  - Compare what happened before screening programmes to what happened after their introduction?
  - Compare the outcomes of those who have been screened with those who have not?
  - Allocate women to screening or not keeping characteristics of groups as balanced as possible and compare the groups, ie a trial?

Breast screening trials

- Recent systematic review (Gotzsche 2011) (more on systematic reviews in week 9)
- Identified 8 trials
- Included 7 trials in analyses
- 600,000 women
- Judged each trial as high or low quality (risk of bias)
- Outcomes:
  - Breast cancer mortality
  - All cancer mortality (risk that radiation exposure may increase other cancers, breast cancer may be cited as cause more often in those screened)
  - Lumpectomy & mastectomy (risk that these may increase in screening group)

Breast screening trials – headline results from the systematic review

Outcomes 13 years after recruitment (screened group compared to non-screened group, rate in unscreened approx 3 per 1000)	Relative risk	95% confidence interval
Breast cancer mortality (3 'high quality 'trials)	0.90	0.79 to 1.02
Breast cancer mortality (4 'low quality' trials)	0.75	0.67 to 0.83 ★
Breast cancer mortality (all 7 trials)	0.81	0.74 to 0.87 ★
All cancer mortality (3 'high quality' trials)	0.99	0.95 to 1.03
Lumpectomies and mastectomies (2 trials only)	1.31	1.22 to 1.42 ★

What do the confidence intervals show? What is the difference between the starred and the un-starred confidence intervals?

More formal treatment of P-values

- Start with null hypothesis
- In this case,
  null hypothesis: no effect of screening
- Do appropriate test (eg chi-squared, log-rank from last week)
- Test gives P-value
- P-value = the probability of getting this result or a more extreme result if the null hypothesis is true
- If P<0.05 we have enough evidence of a difference or an effect to reject the null hypothesis

Results from one trial – Swedish two county trial

Breast cancer mortality (screened group compared to non-screened group)	Relative risk	95% confidence interval	P-value
Local end point committee, 29 years	0.69	0.56 to 0.84	<0.0001
Consensus committee, 29 years	0.73	0.59 to 0.89	0.002
Local end point committee, 10 years	0.74	0.57 to 0.98	<0.05
Consensus committee, 10 years	0.80	0.62 to 1.05	>0.05

What do the confidence intervals tell you?
What do the P-values tell you?
What is the relationship between confidence intervals and P-values?

Understanding controversies about screening

- Different data sources (registries, trials)
- Different biases (eg lead time bias, other biases)
- Different ways of expressing outcomes (life expectancy, mortality)
- Natural history (lead time, overdiagnosis)
- Different presentations/interpretation
- Changes in incidence/prevalence (eg due to lifestyle)
- Changes in treatment (eg improvement)
- Vested interests
- Public opinion
- Is it cost-effective?

(more debate about screening in Health, Society and Illness module week 7)

Wainer's conclusions about breast screening controversies

"In earlier times, physicians had primitive weapons to combat it. The efficacy of those weapons was crucially dependent on early detection. This being the case, the high number of false positives from mammograms could be tolerated. With modern treatments early detection is no longer crucial for success."

What should we tell women?

What would you want to know?
How would you summarise the research information?

(more in seminar)

What about some other screening tests?

- Natural history, accuracy of test, cost, treatments, trends, lead time, risk of overdiagnosis…..etc
- Should we screen?

Wilson and Jungner criteria (WHO 1968)

1. The condition should be an important health problem.
2. There should be a treatment for the condition.
3. Facilities for diagnosis and treatment should be available.
4. There should be a latent stage of the disease.
5. There should be a test or examination for the condition.
6. The test should be acceptable to the population.
7. The natural history of the disease should be adequately understood.
8. There should be an agreed policy on whom to treat.
9. The total cost of finding a case should be economically balanced in relation to medical expenditure as a whole.
10. Case-finding should be a continuous process, not just a "once and for all" project.

Prostate cancer screening in the UK

- Only able to identify the first principle
- No national screening programme
  - Recommendation in 1997
  - Review 2010
  - Due to be reviewed 2013/14
- European screening trial
  - mortality reduced by 20 per cent
    (crude mortality rate 35/100,000)
  - BUT to save one life, 48 additional cases of prostate cancer needed to be treated (Schroder, 2012)

Prostate cancer screening: what's happened in the US?

- Evidence from European trial and other trials
  - Lives saved = minimal
  - Side effects = considerable
  - False positives = considerable
- Thus, in 2011 U.S. Preventive Services Task Force recommended against using the prostate-specific antigen (PSA) test to screen for prostate cancer
- But furore followed, possibly due to (Arkes, 2012):
  - anecdotal or personal evidence stronger influence
  - only partial data considering proportion of those still alive out of those who have been tested
  - not understanding the influence of low prevalence on the efficacy of screening tests even if the screening tests are good

Only considering partial data

- Proportion of those still alive of those tested = 80%
- This sounds good….
- But if the proportion of those still alive of those not tested is also 80%....
- The test does not achieve anything

Conclusions

- Need robust information
- Need to understand all the issues
- Also need clear communication
- Wilson and Jungner criteria important
- Consideration of outcomes of false positives important

Summary

- Incidence, prevalence, data sources
- How good is a test? Sensitivity and specificity
- Positive and negative predictive values, true positives, false positives, true negatives, false negatives
- Two by two tables, effect of prevalence on PPV and NPV
- Lead time & bias, overdiagnosis, cost-effectiveness, trends
- Assessing the effectiveness of screening – trials
- Controversies
- Wilson and Jungner criteria
- Prostate cancer screening

Seminar

- Thinking about communicating information about screening
- More thinking about why screening programmes are and are not introduced

Preparation for next two weeks

For the seminar on screening next week:
Familiarise yourself with the two breast screening leaflets that we are going to look at and compare in the seminar

http://www.cancerscreening.nhs.uk/breastscreen/publications/ia-02.html
http://www.cochrane.dk/screening/mammography-leaflet.pdf

For the following week: Begin to think about the scientific article and accompanying media report you have chosen for your assignment