Mathematics of a False Positive Result

This note describes in mathematical detail the assumptions and steps taken to answer the question: if the result of a medical test is positive for a particular disease what is the probability that an individual will have the disease? In answering this question a context is given for describing the terminology 'false positive'. To illustrate the mathematical steps information was taken from the web about the Prostate Specific Antigen (PSA) test used in detecting Prostate Cancer (PC).

The assumptions are as follows:

A) There are many issues[1] with using raised levels of PSA as a test for prostate cancer, for example 2 out of 3 men will not have prostate cancer with a raised PSA levels[2]. However, for the purposes of the following calculations a test accuracy of 33% will be assumed.

B) As men grow older they are more likely to be diagnosed with prostate cancer. Using the statement taken from the web: "Although only 1 in 10,000 men under age 40 will be diagnosed, the rate shoots up to 1 in 38 for ages 40 to 59, and 1 in 14 for ages 60 to 69"[3]. In the UK the number of men between the age ranges of 40 to 59 is approximately 17 Million[4].

Based on the assumptions above:

The number of men in the population who have prostate cancer is approximately 450,000 (= 17,000,000 / 38). Therefor the number of men who don't have prostate cancer is: 16,550,000.

The number of men who:

1) Have prostate cancer and test positive is 1500,00 (= 450,00 / 3).

2) Have cancer but test negative ('false negative') is 300,00 (= 450,00 * 2 / 3).

3) Don't have cancer but test positive ('false positive') is 11,000,000 (16,550,00 * 2 / 3).

4) Don't have prostate cancer and correctly test negative is approximately 5,550,000 (= 16,550,000 / 3).

The results can be tabulated as follows[5]:

PSA Test	Prostate Cancer	No Cancer	Totals
Positive (+)	150,000	11,000,000	11,150,000
Negative (-)	300,000	5,550,000	5,850,000
Totals	450,000	16,550,000	17,000,000

From the table above and the probabilities can be calculated:

i) The probability that a person picked randomly from the population will have

Prostate Cancer is P(PC) = 450,000 / 17,000,000 = 2.6 % (= 1 / 38).

ii) The probability that the test will successfully detect prostate cancer:

P(PSA +ve) = 150,000 / 450,000 = 33 %.

iii) The probability that a person will have cancer if they test positive is:

P(PC | PSA +ve) = 150,000 / 11,150,000 = 1.3% (odds of approximately 80 to 1)

iv) The probability that the result is 'false positive', in other words the person doesn't have cancer when they test positive:

P(not PC | PSA +ve) = 11,000,000 / 11,1500,000 = 98.7%

The calculation shown in iii above can be carried out using Bayes Theorem[6],:

P(PC | PSA +ve) = P(PSA +ve | PC) * P (PC) / P(PSA +ve),

and using,

P(PSA +ve | PC) = 1500,000 / 450,000,

P(PC) = 450,000 / 17,000,000,

P(PSA +ve) = 11,150,000 / 17,000,000,

therefore,

P(PC | PSA +ve) = 150,000 / 11,150,000 = 1.3%.

References

1. Issues surrounding PSA test and reasons why it currently be used for screening: http://www.nhs.uk/Conditions/Cancer-of-the-prostate/Pages/Prevention.aspx

2. Information taken from the NHS Prostate Patient Information Leaflet: http://www.cancerscreening.nhs.uk/prostate/prostate-patient-info-sheet.pdf

3. Prostate Cancer Foundation – Risk Factors: http://www.pcf.org/site/c.leJRIROrEpH/b.5802027/k.D271/Prostate_Cancer_Risk_Factors.htm

4. Information taken from Wikipedia is 16,978,000: https://en.wikipedia.org/wiki/Demography_of_the_United_Kingdom#Age_structure

5. In the terminology of Venn Diagram's the following table can be developed:

Test	Prostate Cancer	Not Prostate Cancer	Totals
Positive (+)	| PC ∩ PSA +ve |	| not PC ∩ PSA -ve |	| PSA +ve |
Negative (-)	| PC ∩ PSA -ve |	| not PC ∩ PSA +ve |	| PSA –ve |
Totals	| PC |	| not PC |	| Population |

where | A | represents cardinality and ∩ is intersection.

6. Bayes Theorem: https://en.wikipedia.org/wiki/Bayes%27_theorem