Medical statistics, informatics and telemedicine

Base Statistical Concepts in Epidemiology.

Prevalence – Incidence, Risk, Odds. ROC Curves. Likelihood Ratios.

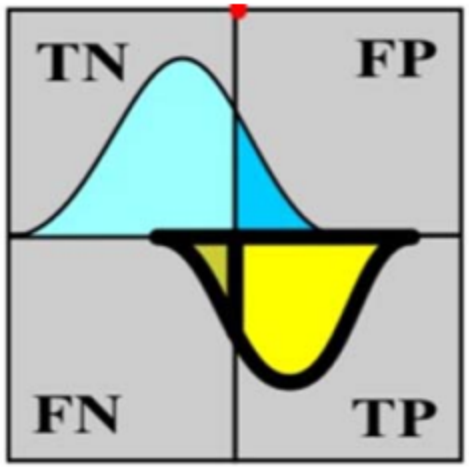
by: Dániel Sándor VERES

19th November 2021

Repetition - Prevalence

Prevalence

= frequency of diseased in examined population
= probability prior to test
= a-priori-probablity



measure of how common the disease is

$$\frac{\text{▽}}{\text{▽△}} = w = \frac{\text{diseased}}{\text{total}} = \frac{TP + FN}{TP + TN + FN + FP} = \frac{de - sp}{se - sp}$$

Incidence

Rate of new cases of disease/year/person in a given population
(e. g. per year, per 10 000 people)

Prevalence - Incidence

Prevalence (point-prevalence)

$$\frac{number\ of\ patients}{population\ size}$$

- Usually in %, if small use /1000(capita), etc.

Incidence (incidence proportion, cumulative incidence)

$$\frac{Number\ of\ \boldsymbol{new\ cases}\ developing\ in\ a\ specified\ period\ of\ time}{population\ size\ \boldsymbol{at\ risk}\ at\ beginning\ of\ the\ interval}$$

- Usually in %-ban, if small use/1000(capita), etc.
- !! We have to give the time interval of obsevation!
  – but this is not the unit!

Prevalence - Incidence

Prevalence (point-prevalence)

$$\frac{number\ of\ patients}{population\ size}$$

- Usually in %, if small use /1000(capita), etc.

Incidence (incidence proportion, cumulative incidence)

$$\frac{Number\ of\ \boldsymbol{new\ cases}\ developing\ in\ a\ specified\ period\ of\ time}{population\ size\ \boldsymbol{at\ risk}\ at\ beginning\ of\ the\ interval}$$

- Usually in %-ban, if small use/1000(capita), etc.
- !! We have to give the time interval of obsevation!
  - – but this is not the unit!

Prevalence - Incidence

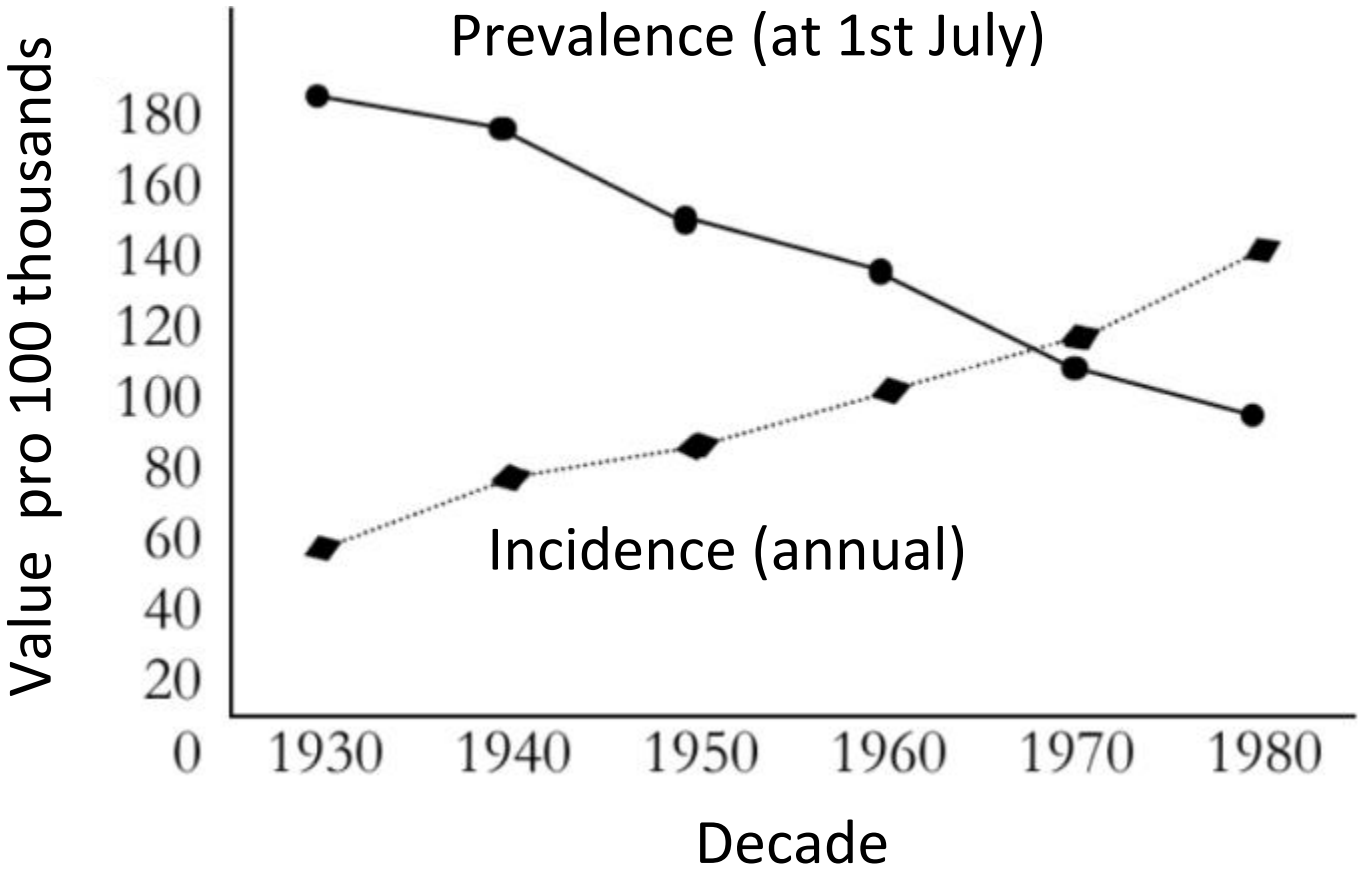
Prevalence

- Shows a „snapshot”
- Probability of „being ill”
- What burden does the disease have at the moment?
- It can be estimated from a cross-sectional examination
- Limited when studying cause-and-effect

Incidence

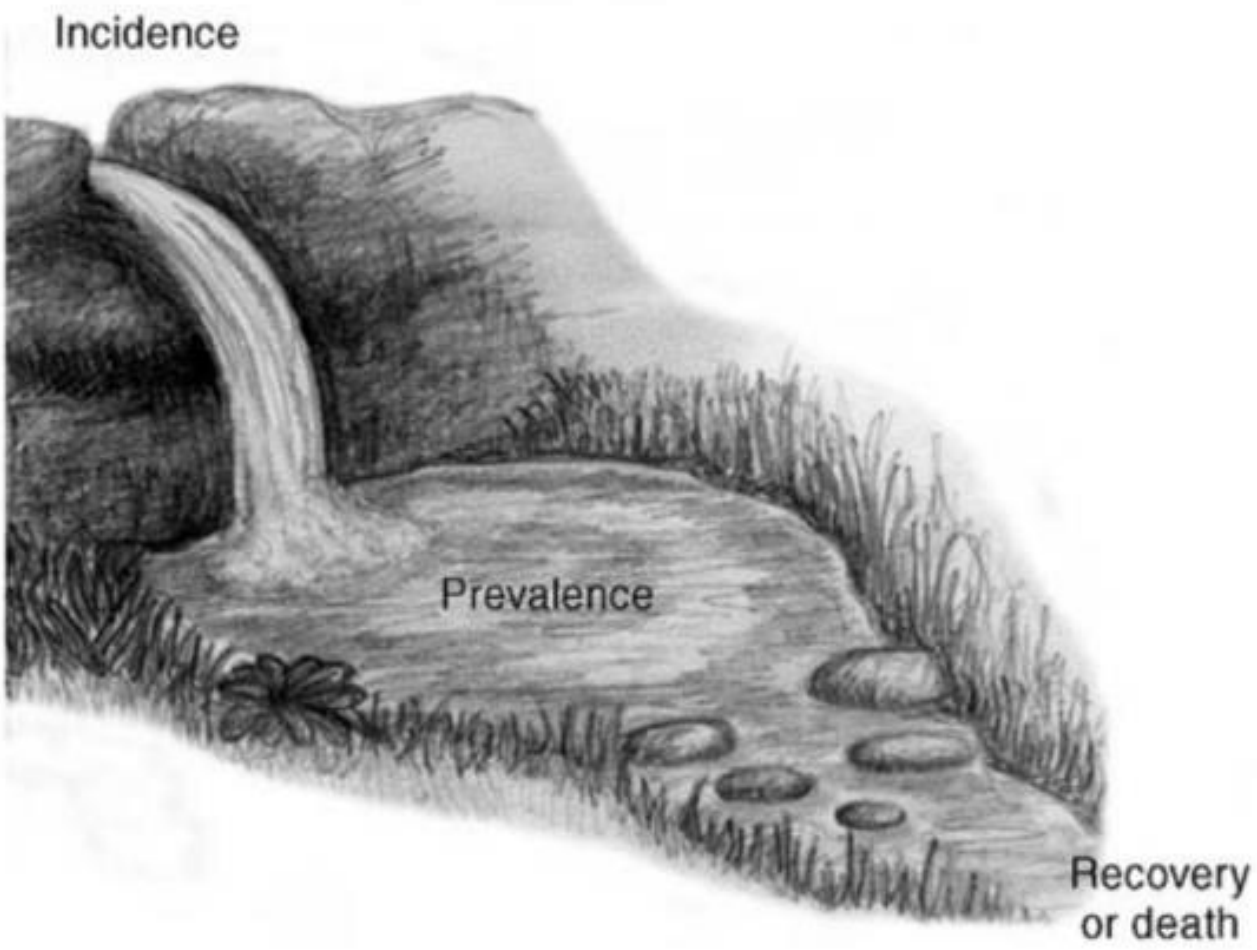
- It shows the “evolution” of a disease
- Probability of “getting sick”
- You can measure the effect of risk factors (can be interpreted as risk)
- It cannot be estimated from a cross-sectional examination
- It is better to study cause-and-effect
- The time interval is important! 1% in 1 year is different from 1% in 20 years

Prevalence-Incidence



Is it possible? HOW?

Lake of the Epidemiologist



Incidence Rate

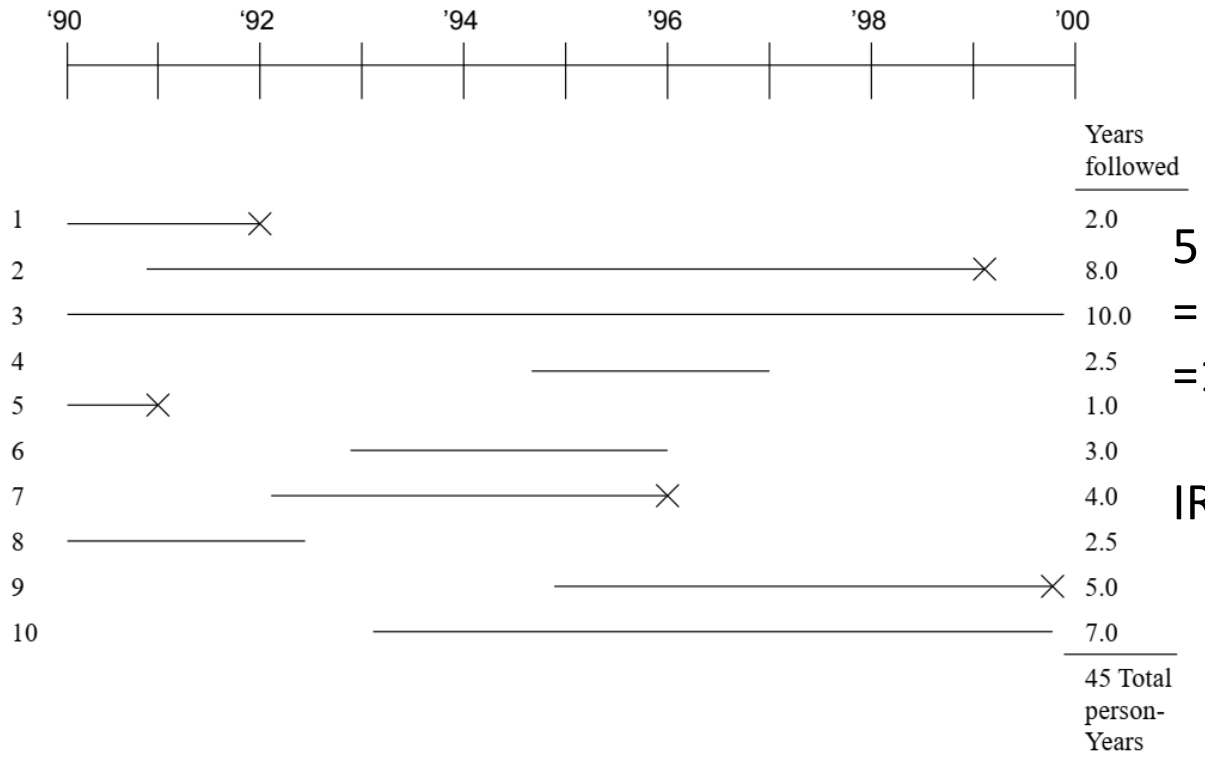
Problematic situation: we could not observe all subject at full length of the observation time interval

- Solution: using time at risk per subject

$$\frac{Number\ of\ \boldsymbol{new\ cases}\ developing\ in\ a\ specified\ period\ of\ time}{Population\ size\ at\ risk\ *\ length\ of\ time\ interval(s)}$$

Incidence Rate

$$\frac{Number\ of\ \boldsymbol{new\ cases}\ developing\ in\ a\ specified\ period\ of\ time}{Population\ size\ at\ risk\ *\ length\ of\ time\ interval(s)}$$



5 cases / 45 person-years
= 0,111 cases / person-years
=11,1 cases / 100 person-years

IR= 11,1 / 100 person-years

- We should use Incidence Rate (IR) if subject could leave or join to the observed population

Risk - Odds

- In epidemiology, we talk about risks and risk factors
- Prevalence and incidence are risks
- Odds is often used (see previous lectures)

Prevalence odds

$$\frac{Number\ of\ patients}{Number\ of\ non-patients} \qquad \frac{Proportion\ of\ patients\ in\ the\ population}{Proportion\ of\ non-patients\ in\ the\ population}$$

Incidence odds

$$\frac{Number\ of\ new\ cases\ developing\ in\ a\ specified\ period\ of\ time}{population\ size\ at\ risk\ at\ \boldsymbol{end}\ of\ the\ interval}$$

$$\frac{Probability\ of\ getting\ sick\ in\ a\ time\ interval\ (\ldots beginning)}{1\ -Probability\ of\ getting\ sick\ in\ a\ time\ interval\ (\ldots beginning)}$$

Relative Measures

- In epidemiology, we talk about risks and risk factors
- Prevalence and incidence are risks
- Odds is often used (see previous lectures)

- Risk *ratios* and odds *ratios* are in common use based on
  a presence/absence of a risk factor (both for prevalence, incidences)
  - Relative measures (as ratios)

$$RR = \frac{Risk\ at\ the\ presence\ of\ a\ risk\ factor}{Risk\ at\ the\ absence\ of\ a\ risk\ factor}$$ (PR(R), IR(R) or (RR), IRR)

$$OR = \frac{Odds\ at\ the\ presence\ of\ a\ risk\ factor}{Odds\ at\ the\ absence\ of\ a\ risk\ factor}$$ (PRR, IR (RR), IRR)

- We could calculate risk and odds *differences* too (RD, OD)
  - Absolute measures (as differences)

Which one – RR or OR?

	Deceased	Survived
Case	25	75
Control	50	50

- What is the risk ratio of death (case versus control group)?
- What is the risk ratio of survive?
- What are the odds ratios?

Which one – RR or OR?

	Deceased	Survived
Case	25	75
Control	50	50

$$RR_{deceased} = \frac{R_{deceased|case}}{R_{deceased|control}} = \frac{\frac{25}{25+75}}{\frac{50}{50+50}} = \frac{1}{2}$$

$$RR_{survived} = \frac{R_{survived|case}}{R_{survived|control}} = \frac{\frac{75}{25+75}}{\frac{50}{50+50}} = \frac{3}{2}$$

$$OR_{deceased} = \frac{O_{deceased|case}}{O_{deceased|control}} = \frac{\frac{25}{75}}{\frac{50}{50}} = \frac{1}{3}$$

$$OR_{survived} = \frac{O_{survived|case}}{O_{survived|control}} = \frac{\frac{75}{25}}{\frac{50}{50}} = \frac{3}{1}$$

Which one – RR or OR?

	Deceased	Survived
Case	25	75
Control	50	50

$$RR_{deceased} = \frac{R_{deceased|case}}{R_{deceased|control}} = \frac{\frac{25}{25+75}}{\frac{50}{50+50}} = \frac{1}{2}$$

$$RR_{survived} = \frac{R_{survived|case}}{R_{survived|control}} = \frac{\frac{75}{25+75}}{\frac{50}{50+50}} = \frac{3}{2}$$

- Non symmetric
- Easier to interpret (ratio of probabilities)
- [better for some calculation]

$$OR_{deceased} = \frac{O_{deceased|case}}{O_{deceased|control}} = \frac{\frac{25}{75}}{\frac{50}{50}} = \frac{1}{3}$$

$$OR_{survived} = \frac{O_{survived|case}}{O_{survived|control}} = \frac{\frac{75}{25}}{\frac{50}{50}} = \frac{3}{1}$$

- Symmetric (benefit?)
- Harder to interpret (ratio of odds)
- [better for some calculation]
  - If deceased/disease is rare, OR~RR

Relative measures – meaning?

- Mammography reduces the risk of breast cancer by 25%.

TRUE / FALSE ?: 250 fewer than 1,000 women die of breast cancer

Relative measures – meaning?

- Mammography reduces the risk of breast cancer by 25%.

TRUE / FALSE ?: 250 fewer than 1,000 women die of breast cancer

risks. In a study of 150 gynaecologists, one-third did not understand the meaning of a 25% risk reduction created by mammography screening. Most of them believed that, if all women were screened, 25% or 250 fewer women out of every 1000 would die of breast cancer, although the best evidence-based estimate is about 1 in 1000.[2,4]

Absolute or relative measures?

You have a 10% chance of dying without treatment.

Which medication are you prefer?

- Drug A: reduces the previous risk by 80%
- Drug B: save 8 by treating 100 people
- It doesn't matter
- I don't know

Absolute or relative measures?

You have a 10% chance of dying without treatment.

Which medication are you prefer?

- Drug A: reduces the previous risk by 80%
- Drug B: save 8 by treating 100 people
- It doesn't matter
- I don't know

Medication A: If you take this medication it will decrease your risk of dying by *80% (four fifths)* over the next year.

Medication B: If 100 people with the disease, like you, take this medication *8* deaths can be prevented over the next year.

Question: Which medication do you want? (Circle your answer.)

(1) Medication A
(2) Medication B
(3) Either Medication A or B
(4) Can't decide

Percentage of Respondents ($n = 470$) Choosing Each Possible Answer for Situation One

Answer	Percentage
Medication A (relative benefit)	56.8
Medication B (absolute benefit)	14.7
Medication A or B	15.5
Can't decide	13.0

Absolute or relative measures?

- 50% more people came to the lecture today than last week!
- Of course, today 2 and last week 1 out of 300
  - It doesn't matter
  - I don't know

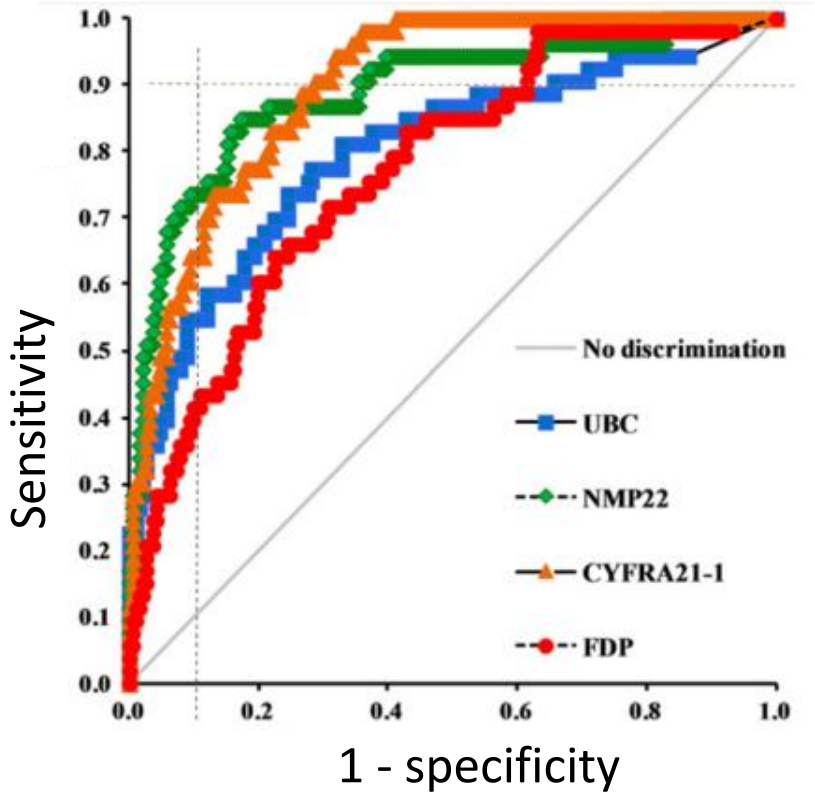
Usually:

- In research, when comparing “different things”: use a relative measure
- For patient, physician (clinical status): use an absolute measure

More diagnostic tests / symptoms

ROC Curves

Comparing tests (for the same illness)?



- WHAT IS IMPORTANT! What is the aim?
  according to the aim use sensitivity, specificity maximum
- Using AUC, YUDEN-index could be misleading

Human Thinking vs, Statistics…?

Our patient's mother turned out to have hereditary breast cancer.
At this situation, such breast cancer is found in the descendants with a **3% probability (prevalence)**.
We perform an MRI scan on our patient, which has a sensitivity of **93% and a specificity of 90%** for this disease.
The test gives a "positive" result.

What do you think, what is the probability of braest cancer in our patient?

O a. < 25%
O b. 25–50%
O c. 50–75%
O d. > 75%

Human Thinking vs, Statistics…?

2. We estimated the probability of the tumor to **20%** based on the patient's symptoms before the test (history, other tests). If we get a “positive” result based on the MR image now, how much is the probability of the tumor? (**93% sensitivity and 90% specificity**)

What do you think, what is the probability in this case?

O a. < 25%
O b. 25–50%
O c. 50–75%
O d. > 75%

A Sequence of Multiple Symptoms, Tests

Example based on Pelkowski et al.

Celiac disease

Prevalence: 1%

if diabetic person: 5%

if a relative (parents, siblings) has celiac disease: 10%

Symptom or test	**Sensitivity (%)**	**Specificity (%)**
Flatulance	76	43
Weight loss	49	57
Loss of appetite	20	81
Diarrhea	71	21
Nausea	20	74
Abdominal pain	37	30
IgA test	96	94

A Sequence of Multiple Symptoms, Tests

Prevalence: 1%

if diabetic person: 5%

if a relative (parents, siblings) has celiac disease: 10%

Symptom or test	**Sensitivity (%)**	**Specificity (%)**
Flatulance	76	43
Weight loss	49	57
Loss of appetite	20	81
Diarrhea	71	21
Nausea	20	74
Abdominal pain	37	30
IgA test	96	94

- What is the odds for celiac disase for a patient with diabetes, flatulance and diarrhea?
- How the odds cahanges if no abdominal pain?
- And if the IgA test is positive?

LR Ratios

How to calculate the final (poterior) probability or odds based on a new test or symptom?
Using **Likelihood-ratios**

$$LR_+ = \frac{Se}{1 - Sp}$$

$$LR_- = \frac{1 - Se}{Sp}$$

LR ratios gives

how much the +/- diagnostic test result increases the odds.

Knowing the prevalence of the disease (Opre):

$$O_{post+} = O_{pre} \cdot LR_+$$

$$O_{post-} = O_{pre} \cdot LR_-$$

LR Ratios

Using 2 tests and both are positive:

$$O_{post++} = O_{pre} \cdot LR_{1+} \cdot LR_{2+}$$

Using 2 tests and one is positive, other one is negative:

$$O_{post+-} = O_{pre} \cdot LR_{1+} \cdot LR_{2-}$$

LR Ratios

Symptom or test	Sensitivity (%)	Specificity (%)	LR+	LR-
Flatulance	76	43	1,33	0,56
Weight loss	49	57	1,14	0,89
Loss of appetite	20	81	1,05	0,99
Diarrhea	71	21	0,90	1,38
Nausea	20	74	0,77	1,08
Abdominal pain	37	30	0,53	2,10
IgA test	96	94	16	0,04

What is the odds for celiac disase for a patient with diabetes, flatulance and diarrhea?

$$\frac{5\%}{95\%} \times 1{,}33 \times 0.9 = 0.063$$

LR Ratios – Example Calculations

Clinical Presentation of Celiac Disease

SYMPTOM	SENSITIVITY (%)	SPECIFICITY (%)	LR+	LR–
Symptoms since childhood	35	89	3.18	0.73
Flatulence/gas	76	43	1.33	0.56
Weight loss	49	57	1.14	0.89
Loss of appetite	20	81	1.05	0.99
Diarrhea	71	21	0.90	1.38
Naus				
Abdo				

Scandinavian Journal of Primary Health Care, 2009; 27: 70–73

ORIGINAL ARTICLE

Sputum colour for diagnosis of a bacterial infection in patients with acute cough

ATTILA ALTINER[1], STEFAN WILM[3], WALTER DÄUBENER[2], CHRISTIANE BORMANN[1], MICHAEL PENTZEK[1], HEINZ-HARALD ABHOLZ[1] & MARTIN SCHERER[4]

We found a bacterial infection significantly more often in yellowish or greenish sputum samples (Pearson's $\chi^2(1) = 6.32$, $p = 0.012$). The sensitivity of and yellowish or greenish sputum used as a test for a bacterial infection was 0.79 (95% CI 0.63–0.94); the specificity was 0.46 (95% CI 0.038–0.53). The positive likelihood ratio (+LR) was 1.46 (95% CI 1.17–1.85) indicating a minimal increase in the likelihood of a bacterial infection for discoloured sputum samples. The positive predictive value (PPV) was 0.16 (95% CI 0.13–0.18).

- See other calculations on the practice