1. What is SIR? standardized incidence ratio

## Interquartile range IQR? What is an Interquartile Range Used For?

Best video: <https://www.youtube.com/watch?v=R6VDj7pEG30>   
The IQR is used to measure how spread out the data points in a set are from the [mean](http://www.statisticshowto.com/mean/)of the data set. The higher the IQR, the more spread out the data points; in contrast, the smaller the IQR, the more bunched up the data points are around the mean. The IQR range is one of many measurements used to measure how spread out the data points in a data set are. It is best used with other measurements such as the median and total range to build a complete picture of a data set’s tendency to cluster around its mean.

**-Confounding variables** (aka third **variables**) are **variables** that the researcher failed to control, or eliminate, damaging the internal validity of an experiment.

[Back to Top](http://www.statisticshowto.com/probability-and-statistics/interquartile-range/#top)

1. <https://stats.stackexchange.com/questions/263318/how-did-researchers-calculate-the-hazard-ratio>
2. Kaplan-Meier find out
3. Medijan je sredina, tak da poredamo sve po velicini I uzmemo sredinu za koju je 50% elementa vece od te sredine I ostalih 50% je manje od te sredine
4. Mod je najcesci podataka u skupu, podatak sa najvise duplikata
5. The **Kaplan**–**Meier** estimator, also known as the product limit estimator, is a non-parametric statistic used to estimate the survival function from lifetime data. In medical research, it is often used to measure the fraction of patients living for a certain amount of time after treatment.
6. Probability!= odds
7. In the cohort study you always start with the exposure and you are looking for the outcome.
8. Odds-izgledi na hrv, can be greater then one
9. Probablity-vjerojatnost
10. Case control is doing the reverse thing
11. Cohort studies are looking for the outcome
12. <https://www.youtube.com/watch?v=hOtoV2Kjb0o>

Many reports in the media about the benefits of treatments present risk results as relative risk reductions rather than absolute risk reductions. This often makes the treatments seem better than they actually are. This leaflet tries to explain the difference between absolute and relative risk. This may enable you to make more informed decisions about whether to take a treatment or not.

What are absolute and relative risks?

**Absolute risk** of a disease is your risk of developing the disease over a time period. We all have absolute risks of developing various diseases such as heart disease, cancer, stroke, etc. The same absolute risk can be expressed in different ways. For example, say you have a 1 in 10 risk of developing a certain disease in your life. This can also be said to be a 10% risk, or a 0.1 risk - depending on whether you use percentages or decimals.

**Relative risk** is used to compare the risk in two different groups of people. For example, the groups could be smokers and non-smokers. All sorts of groups are compared to others in medical research to see if belonging to a group increases or decreases your risk of developing certain diseases. For example, research has shown that smokers have a higher risk of developing heart disease compared to (relative to) non-smokers. Relative risk is usually reported as a percentage (i.e. 10% more likely) but you’ll also see it written as “x times more likely” (i.e. ten times more likely). Although relative risk does provide some information about risk, it doesn’t say anything about the actual odds of something happening; on the other hand, absolute risk does.

A couple of examples may illustrate this better:

**An example when talking about risks of disease**

Say the absolute risk of developing a disease is 4 in 100 in non-smokers. Say the relative risk of the disease is increased by 50% in smokers. The 50% relates to the 4 - so the absolute increase in the risk is 50% of 4, which is 2. So, the absolute risk of smokers developing this disease is 6 in 100.

**An example when talking about treatments**

Say men have a 2 in 20 risk of developing a certain disease by the time they reach the age of 60. Then, say research shows that a new treatment reduces the relative risk of getting this disease by 50%. The 50% is the relative risk reduction, and is referring to the effect on the 2. 50% of 2 is 1. So this means that the absolute risk is reduced from 2 in 20, to 1 in 20.

Number needed to treat (NNT)

A figure which is often quoted in medical research is the NNT. This is the number of people who need to take the treatment for one person to benefit from the treatment.

For example, say a pharmaceutical company reported that medicine X reduced the relative risk of developing a certain disease by 25%. If the absolute risk of developing the disease was 4 in 100 then this 25% reduction in relative risk would reduce the absolute risk to 3 in 100.

However, this can be looked at another way. If 100 people do not take the medicine, then 4 in those 100 people will get the disease. If 100 people do take the medicine, then only 3 in those 100 people will get the disease. Therefore, 100 people need to take the treatment for one person to benefit and not get the disease. So, in this example, the NNT is 100.

A quick way of obtaining the NNT for a treatment is to divide 100 by the absolute reduction in percentage points in risk when taking the medicine. Here is another quick example. Say the absolute risk of developing complications from a certain disease is 4 in 20. Say a medicine reduces the relative risk of getting these complications by 50%. This reduces the absolute risk from 4 in 20, to 2 in 20. In percentage terms, 4 in 20 is 20%, and, 2 in 20 is 10%. Therefore, the reduction in absolute risk in taking this medicine is from 20% to 10% - a reduction of 10 percentage points. The NNT would be 100 divided by 10. That is, 10 people would need to take the medicine for one to benefit.

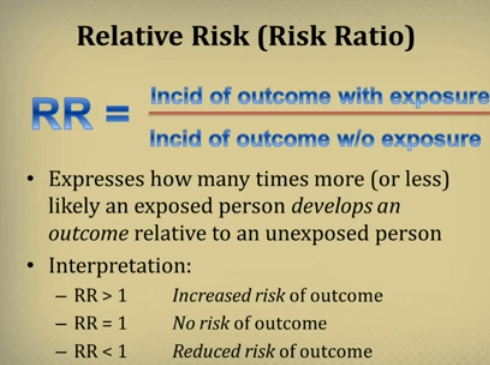
Helping to decide about taking a treatment

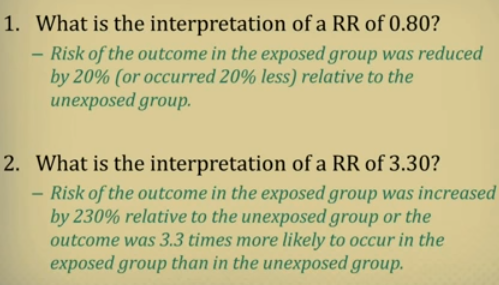
The decision on whether to take a treatment needs to balance various things, such as:

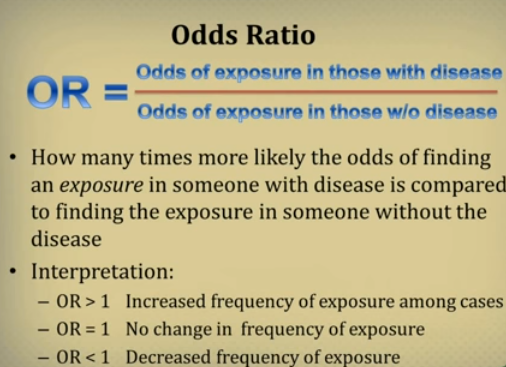
* What is the absolute risk of getting the disease to start with?
* How serious is the disease anyway?
* How much is the absolute risk reduced with treatment?
* What are the risks or side-effects in taking the treatment?
* How much does the treatment cost? Is it worth it to an individual if the individual is paying, or is it worth it to the country if the government is paying?

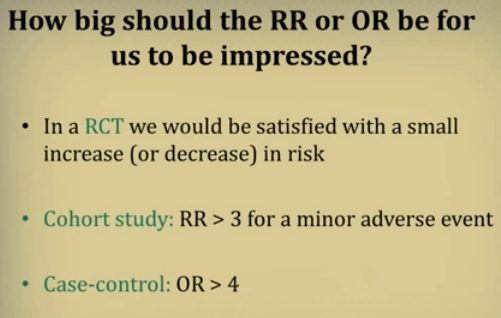
**What is a cohort study:** An example of an epidemiological question that can be answered using a cohort study is: does exposure to X (say, smoking) associate with outcome Y (say, lung cancer)? Such a study would recruit a group of smokers and a group of non-smokers (the unexposed group) and follow them for a set period of time and note differences in the incidence of lung cancer between the groups at the end of this time.

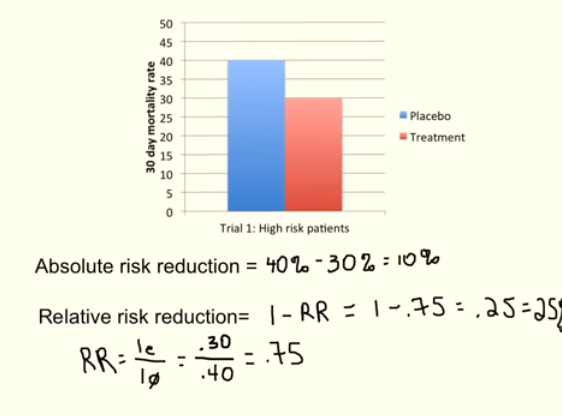
**RELATIVE RISK:**









ODDS je zapravo kad oduzimanje bolesnih ljudi i gledanje na iskljucivo zdrave

ODDS EXPLAINED:

60 PERSON

13 HAVE LUNG CANCER

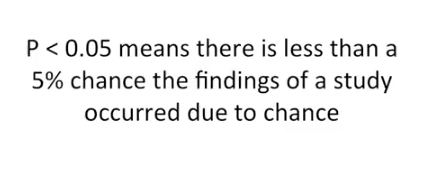
-----------------------

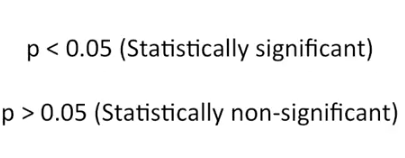
ljude

na nacin da tih 13 u ovom slucaju podjelis sa zdravim ljudima

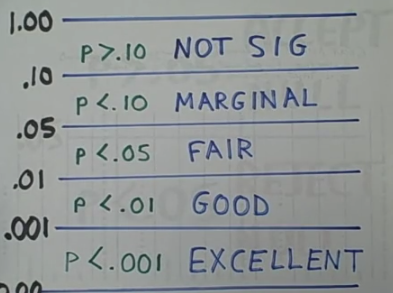
13/47 VJEROJATNOST DA CE NEKO OD ZDRAVIH DOBITI LUNG CANCER JE TOLKO A TO JE ODDS

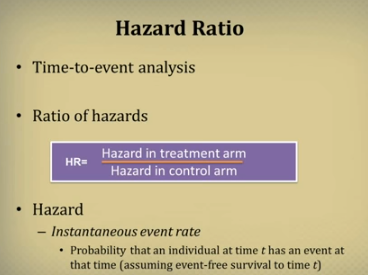
**PIC1 p value**

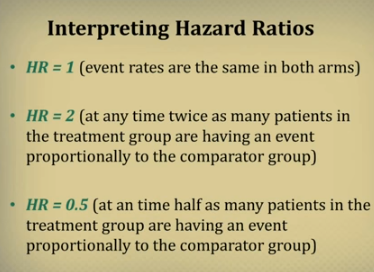




**Rejecting or not rejecting null hypothesis**







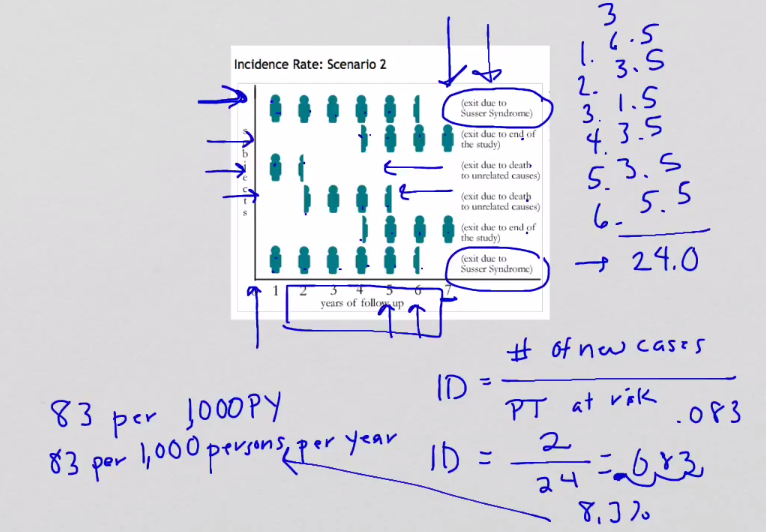
* **Incidence != prevalence**
* **Prevalence** in [epidemiology](https://en.wikipedia.org/wiki/Epidemiology) is the proportion of disease found to have been affecting a particular population (typically a disease or a risk factor such as smoking or seat-belt use)
* **Incidence** in [epidemiology](https://en.wikipedia.org/wiki/Epidemiology) is a measure of the **probability** of occurrence of a given [medical condition](https://en.wikipedia.org/wiki/Medical_condition) in a population within a specified period of time. Although sometimes loosely expressed simply as the number of new cases during some time period, it is better expressed as a proportion or a rate[[1]](https://en.wikipedia.org/wiki/Incidence_(epidemiology)#cite_note-1) with a [denominator](https://en.wikipedia.org/wiki/Denominator_data).

Incidence is number of new cases in a particular population  
Prevalence is number of both old and new cases in a population  
Both measures are useful for different reason inEpidemiology  
  
Total burden of a disease at one point of time is calculated as Prevalence of that disease  
  
whereas Rate of particular disease is calculated as Incidence

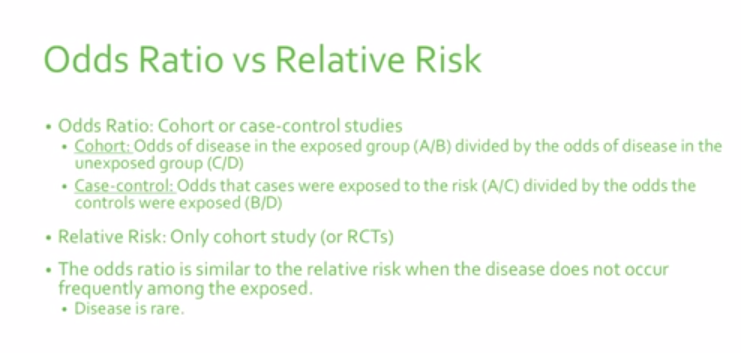
In reality, there were 523 new AIDS cases reported in MA in 2004, and the population was about 5.7 million. So, the cumulative incidence was about 9.2 per 100,000 people during 2004. Note that the denominator is just an estimate based on the last census. In reality, people were being added to and subtracted from the population continually as a result of births, deaths, moving into the city, and moving out. We also didn't take into account exactly when they developed AIDS, although we probably don't care whether they developed it earlier or later within a one year period. Nevertheless, this cumulative incidence is a useful number, and it is relatively easy to get the information we need to calculate it.

**Best explanation for person years:**

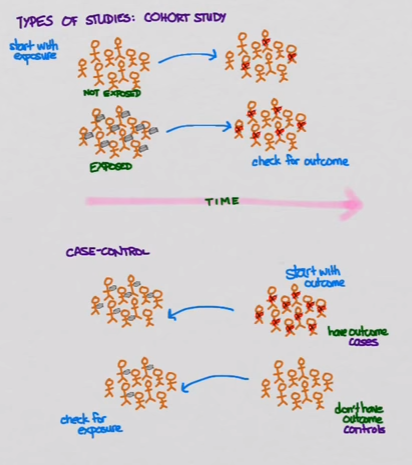
[**https://www.youtube.com/watch?v=YfzopZ\_6KYM**](https://www.youtube.com/watch?v=YfzopZ_6KYM)

****

**Cohort studies** are a type of medical research used to investigate the causes of disease, establishing links between risk factors and health outcomes. **Cohort studies** are usually forward-looking - that is, they are "prospective" **studies**, or planned in advance and carried out over a future period of time.



**Cohort** vs **case study**, cohort study has relative risk case study just odds no relative risk can be calculated



**Incredibly good explanation on median vs mean**

Statistics does not provide a good answer to this question, in my opinion. A mean can be relevant in mortality studies for example, but ages are not as easy to measure as you might think. Older people, illiterate people, and people in some third-world countries tend to round their ages to a multiple of 5 or 10, for instance.

The median is more resistant to such errors than the mean. Moreover, median ages are typically 20 – 40, but people can live to 100 and more (an increasing and noticeable proportion of the population of modern countries now lives beyond 100). People of such age have 1.5 to 4 times the influence on the mean than they do on the median compared to very young people. Thus, the median is a bit more up-to-date statistic concerning a country's age distribution and is a little more independent of mortality rates and life expectancy than the mean is.

Finally, the median gives us a slightly better picture of what the age distribution itself looks like: when you see a median of 35, for example, you know that half the population is older than 35 and you can infer some things about birth rates, ages of parents, and so on; but if the *mean* is 35, you can't say as much, because that 35 could be influenced by a large population bulge at age 70, for example, or perhaps a population gap in some age range due to an old war or epidemic.

Thus, for *demographic,* not *statistical,* reasons, a median appears more worthy of the role of an omnibus value for summarizing the ages of relatively large populations of people.

## Why Use Incidence Rates?

1. To calculate from [population](https://www.ctspedia.org/do/view/CTSpedia/IncidenceRatePopn) based disease registries.
2. To compare [disease incidence in a cohort](https://www.ctspedia.org/do/view/CTSpedia/IncidenceRateCohortPopn) with rate from a general population.
3. To compare [incidence from time-varying exposure](https://www.ctspedia.org/do/view/CTSpedia/IncidenceRateExposure) in persons while exposed and not exposed.

**What are Patient-Years?**

A participant at one of our recent conferences asked a good question—“What are patient-years?”

“Person-years” is a statistic for expressing incidence rates—it is the summing of the results of events divided by time. In many studies, the length of exposure to the treatment is different for different subjects, and the patient-year statistic is one way of dealing with this issue.

The calculation of events per patient-year(s) is the number of incident cases divided by the amount of person-time at risk. The calculation can be accomplished by adding the number of patients in the group and multiplying that number times the years that patients are in a study in order to calculate the patient-years (denominator). Then divide the number of events (numerator) by the denominator.

* Example: 100 patients are followed for 2 years. In this case, there are 200 patient-years of follow-up.
* If there were 8 myocardial infarctions in the group, the rate would be 8 MIs per 200 patient years or 4 MIs per 100 patient-years.

The rate can be expressed in various ways, e.g., per 100, 1,000, 100,000, or 1 million patient-years. In some cases, authors report the average follow-up period as the mean and others use the median, which may result in some variation in results between studies.

Another example: Assume we have a study reporting one event at 1 year and one event at 4 years, but no events at year 2 and 3. This same information can be expressed as 2 events/10 (1+2+3+4=10) years or an event rate of 0.2 per person-year.

An important issue is that frequently the timeframe for observation in studies reporting patient-years does not match the timeframe stated in the study. Brian Alper of Dynamed explains it this way: “If I observed a million people for 5 minutes each and nobody died, any conclusion about mortality over 1 year would be meaningless. This problem occurs whether or not we translate our outcome into a patient-years measure. The key in critical appraisal is to catch the discrepancy between timeframe of observation and timeframe of conclusion and not let the use of ‘patient-years’ mistranslate between the two or represent an inappropriate extrapolation.”[1]

## HR VS RR

Hazard ratio and relative risk   
( risk ratio ) have similar interpretation.  The only difference is when these two are used.  
Hazard ratio gives you instantaneous risk at a particular time.  
and relative risk gives cumulative risk over a time span.   
I hope this explains the whole thing.  
  
**Example:**  
Consider a clinical trial which compares survival of patients with drug  A and drug B.  
  
Relative risk calculated is 2.0  
Hazard ratio calculated is also 2.0. (Normally there is very tiny difference in both the estimates)  
  
Then we can say that,  
***for relative risk:* at the  end of the study without considering time factor, risk of dying with drug A is twice as compared with drug B**  
  
***for hazard ratio:* risk of dying with drug A is twice as compared with drug B at any fixed point in time**