Absolute Risk Reduction (ARR)
(or Risk Difference [RD] or Attributable Risk)

The difference in risks of an outcome between 2 experimental groups. Usually calculated as the difference between the unexposed or control event rate (CER) and the treated or experimental event rate (EER). Sometimes the risk difference is between 2 experimental groups.

\[
\text{ARR} = \text{CER} - \text{EER}
\]

Or

\[
\text{ARR} = \text{EER}_1 - \text{EER}_2
\]

Accuracy

Truthfulness of results or measurements. Requires comparison to known “truth”. Also referred to as validity.

Bias

Systematic error in study design which may skew the results leading to a deviation from the truth. A non-inclusive list of specific types can include:

- Interviewer bias – error introduced by an interviewer's conscious or subconscious gathering of selective data.
- Lead-time bias – mistakenly attributing increased survival of patients to a screening intervention when longer survival is only a reflection of earlier detection in the preclinical phase of disease.
- Recall bias – error due to differences in accuracy or completeness of recall to memory of past events or experiences. Particularly relevant to case control studies (CCS).
- Referral bias – the proportion of more severe or unusual cases tends to be artificially higher at tertiary care centers.
- Selection bias – an error in patient assignment between groups that permits a confounding variable to arise from the study design rather than by chance alone.
- Volunteer bias – people who choose to enroll in clinical research may be systematically different (e.g. healthier, or more motivated) from your patients.
- Workup bias – when the decision to conduct the confirmatory or reference standard test is influenced by the result of the diagnostic test under study (a.k.a verification bias).

Blinded or Masked

Blinded studies purposely deny access to information in order to keep that information from influencing some measurement, observation, or process (therefore blinding reduces information bias). “Double-blinded” refers to the fact that neither the study subject nor the study staff are aware of which group or intervention the subject has been assigned.

Cochrane Collaboration

This international group, named for Archie Cochrane, is a unique initiative in the evaluation of healthcare interventions. The Collaboration works to prepare, disseminate, and continuously update systematic reviews of controlled trials for specific patient problems.

Co-intervention

Interventions other than the treatment under study. Particularly relevant
to therapy RCTs’ - readers should assess whether co-interventions were differentially applied to the treatment and control groups.

Concealment

A fine point associated with randomization that is very important. Ideally, you want to be reassured that the randomization schedule of patients was concealed from the clinicians who entered patients into the trial. Thus the clinician will be unaware of which treatment the next patient will receive and therefore cannot consciously – or subconsciously – distort the balance between the groups. If randomization wasn’t concealed, patients with better prognoses may tend to be preferentially enrolled in the active treatment arm resulting in exaggeration of the apparent benefit of therapy (or even falsely concluding that treatment is efficacious). Note that concealment therefore reduces the possibility of selection bias (compare and contrast with blinding).

Confidence Interval (CI)

Clinical research provides a ‘point estimate’ of effect from a sample of patients; CIs express the degree of uncertainty or imprecision regarding this point estimate. CI represents a range of values consistent with the experimental data. In other words, CIs provide us with the ‘neighborhood’ within which the true (underlying and unknown) value is likely to reside. CIs and its associated point estimate help us make inferences about the underlying population.

The commonly used 95% CI can be defined as the range of values within which we can be 95% sure that the true underlying value lies. This implies that if you were to repeat a study 100 times, 95 of the 100 CIs generated from these “trials” would be expected to include the true underlying value. The 95% CI estimates the sampling variation by adding and subtracting 2 (or 1.96 to be exact) standard errors from the point estimate. The width of the CI is affected by inherent variability of the characteristic being measured, and the study sample size - thus the larger the sample size the narrower (i.e., more precise) the CI. Finally, a useful rule to remember is that values outside of a 95% CI are statistically significantly different (at P < 0.05) from the point estimate. Acceptable formal definitions of the 95% CI in Epi-546 and 547 include:

i) The 95% CI includes a set of values which have a 95% probability or chance of including the underlying true value.

ii) The 95% CI is a measure of the precision surrounding the point estimate.

Confounder or Confounding Variable

A factor that distorts the true relationship of the study variable of interest by virtue of being related to both the study variable and the outcome of interest. Confounders are often unequally distributed among the groups being compared. Randomized studies are less likely to have their results distorted by confounders because randomization should result in the equal balance of these factors at baseline.

Diagnosis

The determination of the nature of a disease; a process of more or less accurate guessing.

Differential Diagnosis

A probabilistic listing of potential causes of a patient’s clinical problem which can be ordered in a probabilistic, prognostic, or pragmatic fashion. Used to aid diagnostic decision-making.

Effectiveness

A measurement of benefit resulting from an intervention for a given health problem under conditions of usual practice. This form of evaluation considers both the efficacy of an intervention and its acceptance by those to whom it is offered. It helps answer “does the practice do more good than harm to people to whom it is offered?”
Efficacy
A measure of benefit resulting from an intervention for a given health problem under conditions of ideal practice. It helps answer “does the practice do more good than harm to people who fully comply with the recommendations?” (N.B. It is the job of RCTs’ to measure efficacy)

Event Rate (risk or CIR)
The risk or CIR of an event. Calculated as the proportion of a fixed population who develop the event of interest over a period of time. In a RCT design the terms controlled event rate (CER) and experimental event rate (EER) refer to the risks in the two comparison groups.

Evidence-Based Medicine
The conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine requires integration of individual clinical expertise and patient preferences with the best available external clinical evidence from systematic research.

Generalizability (or external validity)
The extent to which the conclusions derived from a trial (or study) can be used beyond the setting of the trial and the particular people studied in the trial. Are the results applicable to the full population of all patients with this condition? Also referred to as External Validity. {See ‘internal validity’ and also ‘inference’ which deals with individualizing evidence to specific patients }.

Gold Standard Reference Standard
An established method, or a widely accepted method, for determining a diagnosis. It provides a standard to which a new screening or diagnostic test can be compared. Most importantly in articles about diagnostic tests, the gold standard must be explicitly acknowledged and applied independently in a blinded fashion.

Health
A state of optimal physical, mental, and social well being; not merely the absence of disease and infirmity (World Health Organization).

Health Outcome
All possible changes in health status that may occur in an individual or in a defined population or that may be associated with exposure to an intervention.

Heterogeneity
Differences between patients (clinical heterogeneity) or differences in the results of different studies (statistical heterogeneity).

Inception Cohort
A designated group of persons assembled at a common time early in the development of a specific clinical disorder and who are followed thereafter. In assessing articles about prognosis it is critical that the inception cohort is well described in order to permit assessment of the homogeneity of the cohort.

Incidence rate
Number of new cases of disease occurring during a specified period of time; expressed either as a percentage (or proportion) of the number of people at risk (i.e., cumulative incidence rate [CIR]) or the number of new cases occurring per person time (i.e., incidence density rate - IDR).

Inference
To arrive at a conclusion. The act of taking information from published experience and individualizing to specific patients. The hierarchy and quality of available evidence significantly influence the strength of inference.

Intention-to-Treat Analysis
Analyzing patient outcomes based on which group they were randomized into regardless of whether they actually received the planned intervention. This analysis preserves the power of randomization, thus maintaining that important unknown factors that influence outcome are likely equally distributed in each comparison group. It is the most conservative but valid analytical approach for a RCT (compared to ‘as
Internal validity

The degree to which inferences drawn from a specific study are accurate. Internal validity requires a careful assessment of the study’s methodology to determine whether the observed findings are accurate. Internal validity implies that apart from random error the study’s findings cannot be ascribed to a systematic error or bias; in other words the study does not suffer from confounding, selection or information bias to an important degree (judgment is required here). Thus RCT’s have high internal validity. Contrast with external validity or generalizability.

Kappa (K)

A measure of reliability or agreement between two raters for categorical or qualitative data (e.g., two physicians independently reading x-ray films). Kappa adjusts for the agreement that would be expected to occur due to chance alone, and is thus referred to a chance-corrected agreement. Kappa is preferred over other agreement measures such as the overall % agreement which is highly influenced by the prevalence of the condition being evaluated. Kappa ranges from -1 to +1. Values above 0.80 indicate excellent agreement, values 0.6-0.8 indicates substantial agreement, values 0.4-0.6 indicate moderate agreement, and values < 0.4 indicate fair or poor agreement.

Likelihood Ratio

A ratio of likelihoods (or probabilities) for a given test result. The first is the probability that a given test result occurs among people with disease. The second is the probability that the same test result occurs among people without disease. The ratio of these 2 probabilities (or likelihoods) is the LR. It measures the power of a test to change the pre-test into the post-test probability of a disease being present.

This ratio expresses the likelihood that a given test result would be expected to occur in patients with the target disorder compared to the likelihood of that same result in patients without that disorder. The LR for a given test result compares the likelihood of the result occurring in patients with disease to the likelihood of the result occurring in patients without disease. LRs contrast the proportions of patients with and without disease who have a given test result.

Matching

A deliberate process to make the study group and comparison group comparable with respect to factors (or confounders) that are extraneous to the purpose of the investigation but which might interfere with the interpretation of the studies’ findings. For example in case control studies, individual cases may be matched with specific controls on the basis of comparable age, gender, and/or other clinical features.

Median Survival

Length of time that one-half of the study population survives.

Meta-Analysis (MA)

A systematic review (SR) which uses quantitative tools to summarize the results.

Number Needed to Harm (NNH)

The number of patients who would need to be treated over a specific period of time before one adverse side-effect of the treatment will occur. It is the inverse of the absolute risk increase [ARI] (in the context of a RCT, the ARI is calculated as the EER – CER, where the event rates are adverse events, and by implication, adverse events are more common in the intervention, compared to control group).

$$\text{NNH} = \frac{1}{\text{ARI}}$$

Number Needed to Treat (NNT)

The number of patients who need to be treated over a specific period of time to prevent one bad outcome. When discussing NNTs it is important
to specify the treatment, its duration, and the bad outcome being prevented. It is the inverse of the absolute risk reduction (ARR).

\[ \text{NNT} = \frac{1}{\text{ARR}} \]

**Odds**

A ratio of probability of occurrence to non-occurrence of an event

\[ \text{Odds} = \frac{\text{probability}}{1 - \text{probability}} \]

**Odds Ratio (OR)**

The *odds* ratio is most often used in case-control study (CCS) designs to describe the *magnitude or strength* of an association between an exposure and the outcome of interest (i.e. the odds of exposure in cases compared to the odds of exposure in the controls – it is therefore calculated as a ratio of odds). Because the actual underlying disease risks (or CIRs) in the exposed and unexposed groups cannot be calculated in a CCS design, the OR is used to approximate the RR. The OR more closely approximates the RR when the outcome of interest is infrequent or rare (i.e., <10%). As a measure of the strength of association between an exposure and outcome the OR has the same interpretation as the RR. The OR is calculated as the cross-product ratio \(\frac{a \cdot d}{b \cdot c}\) where \(a, b, c,\) and \(d\) represent the respective cell sizes.

**Outcomes**

All possible changes in health status that may occur in following subjects or that may stem from exposure to a causal factor or to a therapeutic intervention.

**P-value**

The probability of obtaining the value of a test statistics at least as large as the one observed, under the assumption that the \(H_0\) is true (i.e. \(P(\text{data}|H_0 \text{ true})\)). The smaller the p-value, the lower your degree of belief is in the null hypothesis being true. From a hypothesis testing standpoint the p-value is used to make a decision based on the available data. {note that increasingly the reporting of confidence intervals (CI) are preferred over p-values because they provide much more useful information – including the precision of the data and their possible clinical significance }.

**Population attributable risk (PAR)**

The incidence of disease in a population that is associated with a risk factor. Calculated from the Attributable risk (or RD) and the prevalence (P) of the risk factor in the population i.e.,

\[ \text{PAR} = \text{Attributable risk} \times P \]

Example: Thromboembolic disease (TED) and oral contraceptives (OC):

Incidence of TED in overall population = 7 per 10,000 person years
Incidence of TED in OC users = 16 per 10,000 person years
Incidence of TED in non-OC users = 4 per 10,000 person years
25% of women of reproductive age take OC

So, \(\text{PAR} = [16 - 4] \times 0.25 = 3\) per 10,000 person years. This represents the excess incidence of TED in the population due to OC use.

**Population attributable risk fraction (PARF) (a.k.a etiologic fraction)**

The fraction of disease in a population that is attributed to exposure to a risk factor. Under the assumption that the risk factor is a cause of the disease, it represents the maximum potential impact on disease incidence if the risk factor was removed. It can be calculated from the PAR as:

\[ \text{PARF} = \frac{\text{PAR}}{\text{Total Disease Incidence}} \]

OR, it can be calculated directly from the RR and P as follows:
PARF = \( \frac{P(RR-1)}{1 + P(RR-1)} \)

Example: Thromboembolic disease (TED) and oral contraceptives (OC):

PARF = PAR/Total Disease Incidence
PARF = 3 per 10,000/7 per 10,000 = 43%

PARF = \( \frac{P(RR-1)}{1 + P(RR-1)} \)
PARF = 0.25(4-1)/ [1 + 0.25(4-1)] = 43%. This represents the fraction of all TED in the population that is due to OC use.

Power
Ability to detect a difference between two experimental groups if one in fact exists (1 – beta).

Precision
A measure of variability in the point estimate as quantified by the confidence interval.

Predictive Value
Positive Predictive Value (PVP) – proportion of people with a positive test who have disease.

Negative Predictive Value (PVN) – proportion of people with a negative test who are free of disease.

Prevalence (P, Prev)
Proportion of persons affected with a particular disease at a specified time. Prevalence rates obtained from high quality studies can inform clinician’s efforts to set anchoring pretest probabilities for their patients. In diagnostic studies, also referred to as prior probability.

Prognosis
The possible outcomes of a disease and the frequency with which they can be expected to occur.

Randomization
Allocation of individuals to groups by a formal chance process such that each patient has an independent, equal chance of selection for the intervention group.

Relative Risk (Risk Ratio)
The relative probability or risk of an event or outcome in one group compared to another. In a RCT, it is calculated as the ratio of risk in the treatment or experimental group (EER) to the risk in the control group (CER), and is a measure of the efficacy (or magnitude) of the treatment effect. In a cohort study, where it is similarly used to express the magnitude or strength of an association, it is calculated as the ratio of risk of disease or death among an exposed population to the risk among the unexposed population.

\[ RR = \frac{EER}{CER} \]

Relative Risk Reduction (RRR)
The percent reduction in an outcome event in the experimental group as compared to the control group. It is the complement of RR or the proportion of risk that’s removed by the intervention. Unlike the ARR the RRR is assumed to be a constant entity – that is, it is assumed not to change from one population (study) to another. It therefore represents a fixed measure of the efficacy of an intervention (contrast with the ARR which is responsive to changes in the baseline event rates).

\[ RRR = 1 – \text{Relative Risk} \]
\[ RRR = CER – EER / CER \times 100 \]
\{Or obviously … \[ RRR = ARR / CER \}\}

Reliability
Refers to consistency or reproducibility of data; a.k.a. repeatability. Referred to as agreement when examining categorical data (see also Kappa). \textit{Intra-rater} reliability refers to the consistency within the same
observer or instrument. Inter-rater reliability refers to the consistency between two observers or instruments. It is important to distinguish reliability from validity.

ROC Curves

Receiver operator characteristic (ROC) curves plot test sensitivity (on the y axis) against 1- specificity (on the x axis) for various cut-points of a continuously distributed diagnostic variable. The curves describe the tradeoff between Se and Sp as the cut point is changed. Test that discriminate well crowd towards the upper north-east corner of the graph. ROC curves can be used to compare the discriminating ability of two or more tests by comparing the area under the curve.

Sensitivity (Se)

The proportion of people with disease who have a positive test or P(T+|D+)

SnNout

When a test with a high sensitivity is negative, it effectively rules out the diagnosis of disease.

Sensitivity Analysis

A test of the stability of conclusions by evaluating the outcome over a range of plausible estimates, value judgments, or assumptions.

Specificity (Sp)

The proportion of people without disease who have a negative test or P(T-|D-).

SpPin

When a test is highly specific, a positive result can rule in the diagnosis.

Standards

Authoritative statements of minimal levels of acceptable performance or results, excellent levels of performance or results, or the range of acceptable performance or results.

Study Designs

1. **Case Series** – A collection or a report of the series of patients with an outcome of interest. No control group is involved.

2. **Case Control Study (CCS)** – Identifies patients who have a condition or outcome of interest (cases) and patients who do not have the condition or outcome (controls). The frequency that subjects are exposed to a risk factor of interest is then compared between the cases and controls. Because of the design of the CCS, disease rates cannot be directly measured (contrast this with the cohort study design). Thus the comparison between cases and controls is actually done by calculating the odds of exposure in cases and controls. The ratio of these 2 odds results in the odds ratio (OR) which is usually a good approximation of the relative risk (RR)

Advantages: it is relatively quick and inexpensive requiring fewer subjects than other study designs. It is often times the only feasible method for investigating very rare disorders or when a long lag time exists between an exposure of interest and development of the outcome/disease of interest. It is also particularly helpful in studies of outbreak investigations where a quick answer followed by a quick response is required. Disadvantages: recall bias, unknown confounding variables, and difficulty selecting appropriate control groups.

3. **Crossover Design** – A method of comparing 2 or more treatments or interventions in which all subjects are switched to the alternate treatment after completion of the first treatment. Typically allocation to the first treatment is by a random process. Since all subjects serve as their own controls, error variance is reduced.

4. **Cross-Sectional Survey** – The observation of a defined population at
a single point in time or during a specific time interval. Exposure and outcome are determined simultaneously. Also referred to as a prevalence survey because this is the only epidemiological frequency measure that can be measured (in other words incidence rates cannot be generated from this design).

5. **Cohort Study** – Involves identification of two groups (cohorts) of patients who are defined according to whether they were exposure to a factor of interest e.g., smokers and non-smokers. The cohorts are then followed over time and the incidence rates for the outcome of interest in each group are measured. The ratio of these incidence rates results in the relative risk (RR) which quantifies the magnitude of association between the factor and outcome (disease). Note that when the follow-up occurs in a forward direction the study is referred to as a prospective cohort. When follow-up is done based on historical information it is referred to as a retrospective cohort.

Advantages: can establish clear temporal relationships between exposure and disease onset. Are able to generate incidence rates.

Disadvantages: control/unexposed groups may be difficult to identify, exposure to a variable may be linked to a hidden confounding variable, blinding is often not possible, randomization is not present. For relatively rare diseases of interest, cohort studies require huge sample sizes and long f/u (hence they are slow and expensive).

6. **N-of-1 Trial** – When an individual patient undergoes pairs of treatment periods organized so that one period involves use of the experimental treatment and the other involves use of a placebo or alternate therapy. Ideally the patient and physician are both blinded, and outcomes are measured. Treatment periods are replicated until patient and clinician are convinced that the treatments are definitely different or definitely not different.

7. **Randomized Controlled Trial** – A group of patients is randomized into an experimental group and into a controlled group. These groups are then followed up and various outcomes of interest are documented. RCT’s are the ultimate standard by which new therapeutic maneuvers are judged. Randomization should result in the equal distribution of both known and unknown confounding variables into each group. An unbiased RCT also requires concealment and where feasible blinding.

Disadvantages: often impractical, limited generalizability, volunteer bias, significant expense, and sometimes ethical difficulties.

8. **Systematic Review** – A formal review of a focused clinical question based on a comprehensive search strategy and structured critical appraisal designed to reduce the likelihood of bias. No quantitative summary is generated however.

9. **Meta-Analysis** – A systematic review which uses quantitative methods to combine the results of several studies into a pooled summary estimate.
Substitute or Surrogate Endpoints

Refer to study outcomes that are not immediately significant in clinical patient care. Substitute endpoints may include rates of biochemical changes (e.g., cholesterol, HbA1C) while clinically significant endpoints are more clearly tied to events that patients and their doctors care about most (e.g., stroke, renal failure, death).

Validity

Truthfulness or believability of study conclusions or the extent to which a test actually measures what it is supposed to measure or accomplishes what it is supposed to accomplish. Simply put, “Does the data really mean what we think it does?” or “Can we believe the results?” A.k.a. accuracy. Validity implies the presence of a gold standard to which data can be compared to. See also internal validity and external validity.

Need online access to these course materials?

Go to: http://learn.chm.msu.edu/epi/

Or visit the Echt homepage and look for the link.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>Attributable risk</td>
</tr>
<tr>
<td>ARI</td>
<td>Absolute risk increase</td>
</tr>
<tr>
<td>ARR</td>
<td>Absolute risk reduction</td>
</tr>
<tr>
<td>CER</td>
<td>Control event rate</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CCS</td>
<td>Case-control study</td>
</tr>
<tr>
<td>CIR</td>
<td>Cumulative incidence rate</td>
</tr>
<tr>
<td>EER</td>
<td>Experimental event rate</td>
</tr>
<tr>
<td>IDR</td>
<td>Incidence density rate</td>
</tr>
<tr>
<td>MA</td>
<td>Meta-analysis</td>
</tr>
<tr>
<td>NNH</td>
<td>Number needed to harm</td>
</tr>
<tr>
<td>NNT</td>
<td>Number needed to treat</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>P (or Prev)</td>
<td>Prevalence</td>
</tr>
<tr>
<td>PAR</td>
<td>Population attributable risk</td>
</tr>
<tr>
<td>PARF</td>
<td>Population attributable risk fraction</td>
</tr>
<tr>
<td>PVP</td>
<td>Predictive value positive</td>
</tr>
<tr>
<td>PVN</td>
<td>Predictive value negative</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>RD</td>
<td>Risk difference</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operator characteristic curve</td>
</tr>
<tr>
<td>RR</td>
<td>Risk ratio or relative risk</td>
</tr>
<tr>
<td>RRR</td>
<td>Relative risk reduction</td>
</tr>
<tr>
<td>Se</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Sp</td>
<td>Specificity</td>
</tr>
<tr>
<td>SR</td>
<td>Systematic review</td>
</tr>
<tr>
<td>TER</td>
<td>Treatment event rate</td>
</tr>
</tbody>
</table>