Foundations in Biostatistics and Epidemiology
Session 7: Epidemiologic Measures of Association
The following provides a summary of the content of this module.

I. Estimating Association Between Exposure and Disease

- Relative Risk (RR)
  - Design: cohort study (typically prospective, participants sampled based on exposure status and followed for development of disease)
  - Calculation:

	Disease (D)	No Disease (ND)	Total
Exposed (E)	a	b	a+b (fixed by sampling design)
Not Exposed (NE)	c	d	c+d (fixed by sampling design)
Total	a+c	b+d	a+b+c+d

RR = [a/(a+b)] / [c/(c+d)] = $Risk_{exposed}$ / $Risk_{not\ exposed}$ = P(D|E) / P(D|NE)

  - Interpretation:
    - RR = 1 → No association
    - RR > 1 → Exposed are at higher risk of outcome than the unexposed
    - RR < 1 → Exposed are at lower risk of outcome than the unexposed
- Incidence Rate Ratio (IRR)
  - Design: prospective cohort study with varying lengths of follow-up (person-time at risk) on participants (participants sampled based on exposure status and followed for development of disease)
  - Calculation:

	Disease (D)	Person-time at Risk
Exposed (E)	D_E	PT_E
Not Exposed (NE)	D_{NE}	PT_{NE}

IRR = [D_E/PT_E] / [D_{NE}/PT_{NE}] = $IR_{exposed}$ / $IR_{not\ exposed}$

  - Interpretation:
    - IRR = 1 → No association
    - IRR > 1 → Exposed have a higher disease incidence rate than the unexposed
    - IRR < 1 → Exposed have a lower disease incidence rate than the unexposed
- Odds Ratio (OR)
  - Design: case-control study (participants sampled based on disease status and exposure information is assessed using participant recall or existing records; could also be calculated for cohort or cross-sectional studies)
  - Calculation of exposure OR for case-control study

	Disease (D)	No Disease (ND)	Total
Exposed (E)	a	b	a+b
Not Exposed (NE)	c	d	c+d
Total	a+c (fixed by sampling design)	b+d (fixed by sampling design)	a+b+c+d

Exposure Odds = P(E)/[1-P(E)] = P(E)/P(NE)
OR = [P(E|D)/P(NE|D)] / [P(E|ND) / P(NE|ND)] = [a/c] / [b/d] = [ad] / [bc] = $ExpOdds_{diseased}$ / $ExpOdds_{not\ diseased}$

  - Interpretation (case-control study):
    - OR = 1 → No association
    - OR > 1 → Diseased participants have a higher odds of exposure compared to non-diseased participants
    - OR < 1 → Diseased participants have a lower odds of exposure compared to non-diseased participants
  - Notes:
    - When calculating the OR for a cohort study (exposure totals are fixed), interpret odds ratio as the disease odds ratio (ratio of the odds of developing disease in exposed persons to the odds of developing disease in non-exposed persons). The calculation is the same: OR=[ad] / [bc]
    - When calculating the OR for a cross-sectional study, interpret odds ratio as the prevalence odds ratio (ratio of the odds of having disease in persons who have the exposure to the odds of having disease in persons who don't have the exposure). The calculation is the same: OR=[ad] / [bc]

d. Prevalence Proportion Ratio (PPR)

- Design: cross-sectional study (only the presence/absence of disease and presence/absence of exposure are measured, do not observe temporal sequence between exposure and disease outcome)
- Calculation:

	Disease (D)	No Disease (ND)	Total
Exposed (E)	a	b	a+b
Not Exposed (NE)	c	d	c+d
Total	a+c	b+d	a+b+c+d

PPR = [a/(a+b)] / [c/(c+d)] = $Prev_{exposed}$ / $Prev_{not\ exposed}$ = P(D|E) / P(D|NE)

- Interpretation:
  - PPR = 1 → No association
  - PPR > 1 → Prevalence of disease in the exposed is higher than the prevalence of disease in the unexposed
  - PPR < 1 → Prevalence of disease in the exposed is lower than the prevalence of disease in the unexposed

II. Estimating Potential for Prevention

a. Attributable Risk (risk difference):

- Design: cohort study (typically prospective, participants sampled based on exposure status and followed for development of disease); assume that association between exposure and disease outcome is causal with all other factors equally distributed between exposed and unexposed
- Calculation:

	Disease (D)	No Disease (ND)	Total
Exposed (E)	a	b	a+b
Not Exposed (NE)	c	d	c+d
Total	a+c	b+d	a+b+c+d

AR = $Risk_{exposed}$ – $Risk_{not\ exposed}$ = [a/(a+b)] – [c/(c+d)]

- Interpretation: amount of disease risk, among exposed, that can be attributed to a specific exposure

b. Attributable Risk Percent

- Design: cohort study (typically prospective, participants sampled based on exposure status and followed for development of disease); assume that association between exposure and disease outcome is causal with all other factors equally distributed between exposed and unexposed
- Calculation:

	Disease (D)	No Disease (ND)	Total
Exposed (E)	a	b	a+b
Not Exposed (NE)	c	d	c+d
Total	a+c	b+d	a+b+c+d

AR% = [$Risk_{exposed}$ – $Risk_{not\ exposed}$] / $Risk_{exposed}$ * 100% = {[a/(a+b)] – [c/(c+d)]} / [a/(a+b)] * 100% = (RR-1)/RR *100%

- Interpretation: Percentage of risk in exposed persons that can be attributed to a specific exposure

c. Population Attributable Risk

- Design: cohort study (typically prospective, participants sampled based on exposure status and followed for development of disease); assume that association between exposure and disease outcome is causal with all other factors equally distributed between exposed and unexposed
- Calculation:

	Disease (D)	No Disease (ND)	Total
Exposed (E)	a	b	a+b
Not Exposed (NE)	c	d	c+d
Total	a+c	b+d	a+b+c+d

PAR = $Risk_{population}$ – $Risk_{not\ exposed}$ = [(a+c)/(a+b+c+d)] – [c/(c+d)]

- Interpretation: amount of disease risk, in the population, that can be attributed to a specific exposure

d. Population Attributable Risk Percent

- Design: cohort study (typically prospective, participants sampled based on exposure status and followed for development of disease); assume that association between exposure and disease outcome is causal with all other factors equally distributed between exposed and unexposed
- Calculation:

	Disease (D)	No Disease (ND)	Total
Exposed (E)	a	b	a+b
Not Exposed (NE)	c	d	c+d
Total	a+c	b+d	a+b+c+d

PAR% = [$Risk_{population}$ – $Risk_{not\ exposed}$] / $Risk_{population}$ * 100% = {[(a+c)/(a+b+c+d)] – [c/(c+d)]} / [(a+c)/(a+b+c+d)] * 100%

- Interpretation: Percentage of risk in the population that can be attributed to a specific exposure

III. Principles of Estimation: confidence intervals

- Point estimates of measures of association, e.g. OR and RR, are often coupled with a 95% confidence interval
- 95% confidence interval:
    - Provides a range of plausible values for the true, population measure of association, e.g., OR or RR
    - Interpretation of 95% confidence interval (LL, UL): we are 95% confident that the true measure of association lies within the interval (LL, UL)
    - The width conveys information about the degree of precision, which is a function of the level of confidence (e.g., 95% or 99%) and variability in our estimate
    - For relative measures, e.g., RR or OR, compare the confidence interval to a value of 1 (i.e., no association between exposure and outcome) to determine if the association is statistically significant. If the interval lies entirely above or below 1 (i.e., does not include 1), the association is statistically significant.
- Factors impacting the width of the confidence interval
    - Increased certainty in our estimation (e.g., consider a 99% confidence interval instead of a 95% confidence interval) $\rightarrow$ wider interval (we want to be more certain in our estimation, so the interval needs to be wider)
    - Increased variability in our estimate $\rightarrow$ wider interval
    - Decreased sample size $\rightarrow$ wider interval
    - Narrower interval indicates increased precision