BE.104 Spring
"Epidemiology: Test Development and Relative Risk"
J. L. Sherley

Agent X $\xrightarrow{\text{? Cause}}$ Δ Health

First, Some definitions

Morbidity = sickness, disease, [toxicity] – Hard to quantify

Mortality = death – Easy to quantify

Annual Death Rate = $\frac{\text{Total deaths during specified 12 months}}{\text{Number of persons in pop. @ middle of the period specified}}$

Age-specific ADR = ADR for a specified age

Age-adjusted ADR = adjusted ADR (indexing for age differences, e.g., death rates in nursing home versus general population)

Incidence = # of **NEW** cases in a population
during a specified period of time / Number of persons exposed to risk of developing disease in same interval (by convention per 100,000 persons) = "Rate of appearance in group at risk"

Prevalence = $\frac{\text{\# of cases in population at a specified time}}{\text{\# of persons present in pop. at same time (By convention per 100,000 persons)}}$ = "Frequency"

[Prevalence α Incidence X duration of disease based on death or cure]
"Chronic disease will be more prevalent"

Tests

To determine these epi-parameters, we need to develop tests to detect and quantify — disease / exposure / death

Develop tests
- Disease measurements (death is not all that easy to quantify when you want death from a specific cause)
- Exposure measurements to X
- Measure X in the environment

Properties of a good test

Simple

Well tolerated < by all involved; subject & tester >

High throughout

Proven versus newly developed

Precise

Accurate: *Sensitive, Specific, Predictive*

Categories of test results

w/D	w/oD	<**Disease (D) Status**
TP	FP	Positive
FN	TN	Negative <**Test**

True positives, TP
False positives, FP
False negatives, FN
True negatives, TN

Sensitivity = $\frac{TP}{TP + FN}$ x 100%; Ideally, = 100% $\Rightarrow$ FN = 0
(All the affected)

Sensitivity = the efficiency at detecting cases

Specificity = $\frac{TN}{TN + FP}$ x 100%; Ideally = 100% $\Rightarrow$ FP = 0
(All the unaffected)

Specificity = efficiency at identifying non-cases

Predictive value (PV): the probability that individuals with a positive (negative) test have the disease (are disease free). How accurate is the test value?

Positive PV = TP/ (TP+FP) ; note: TP+FP = total positive tests. What fraction are true?

Negative PV = TN/ (TN+FN); note: TN+TN = total negative tests. What fraction are true?

Test quality depends on multiple factors

- Degree of difference between cases & controls for a given measure
- Set point for test threshold

Diabetes Example

Sensitivity & Specificity are co-variants with test threshold

In practice, set threshold based on:

1) Disease severity (i.e. don't want to miss things that lead to death or severe injury)
2) Side-effects of response (i.e. don't want to treat normals with potentially harmful treatment)
3) Must consider cost of unnecessary response

What is the point of this example?

Before designing a test to distinguish between affected and non-affected populations we need to know the characteristics of the distribution of values of individuals in the two compared populations.

First case of a "frequency plot" (= distribution).

Risk

Much of modern epidemiology and environmental health science is concern with Risk

Risk- a measure of the likelihood that an event will occur in the future (usually adverse event; but formally it could be positive or neutral)

Since there are no true seers or fortune-tellers, absolute measures of risk are hard to come by.

So, typically Risk is:

1) Past experience-based ("trial & error learning")

   What the risk of holding your hand for 1 minute in a candle flame that you will burn your hand?

   What about on a windy day? Etc.!

2) Seldom individual- population based

3) Often **relative**

   A way to quantify the effect of a presumed risk factor or chemical agent

Odds Ratio- An estimate of relative risk, RR

Consider a disease that occurs within two distinct populations; one known to be exposed to an agent, the other known not to be exposed. How can we estimate the magnitude of any change in risk for the disease that is associated with the exposure to the agent?

	Number of Persons		
Exposure	Affected	Unaffected	Total
Yes	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	a + b + c + d = N

If $\frac{a}{a+c}$ is statistically different than $\frac{b}{b+d}$

$\frac{a}{a+c}$: Fraction of affected who were exposed

$\frac{b}{b+d}$: Fraction of **un**affected who were exposed,

…then we can use the relative risk calculation to quantify the magnitude of the effect of exposure to the agent on the occurrence of the disease.

Ideally: 100% 0%

If these are equivalent, then exposure is not an important determinant of illness.

Affected and unaffected have equal likelihood of exposure.

** Why are these seldom the ideal?
Genes, environment, luck, error

$\frac{a}{a+b}$ = Number new cases who were exposed during period of investigation / Total number exposed during same period (i.e. at risk) = $\text{Incidence}_{\text{exposed}}$

$\frac{c}{c+d}$ = Number new cases who were unexposed during period of investigation / Total number of unexposed during same period = $\text{Incidence}_{\text{unexposed}}$

"Odds Ratio" ≈ $\frac{I_{exposed}}{I_{unexposed}} = \frac{a\,(c + d)}{(a + b)\,c}$ = Relative Risk

$= \frac{ac + ad}{ac + bc}$

$\approx \frac{ad}{bc}$ when $a + c$ <<< $b + d$

total affected *total unaffected*

which is often the case!!!

[Statistical Test: often chi-square for case of RR = 1.0; w/ 95% CI for chi-square distribution]

What level of RR is <u>meaningful?</u>

Given that the calculation is statistically signficant, how can you increase your level of confidence that an estimate of RR is <u>meaningful</u> in terms of public health?

1) Magnitude: consider 1.2 vs 20

   Comfort zone @ RR ≥ 5.0

2) Dose response analysis

Although seeing a dose-response for a suspected toxic factor and relative risk increases confidence for significance of small RR and cause-effect relationship, it is hard to do experimentally because of ethics, unless by <u>reducing</u> exposure.

What else could be done to evaluate dose-response?

<u>Stratification</u>- the "dose response" of epidemics & environmental health scientists

Stratify RR analyses by e.g.,:	Place-	up vs down stream & distance
		up vs down wind & distance
		outdoors vs indoor
	Persons-	workers vs non-workers
		fish eaters vs others
		children vs adults
	Time-	before vs after the plant
		resident vs moved away
		new resident vs old resident

Look for <u>statistically significant associations</u> with RR

One major limitation: Decrease sample size for group stratification (More later on this!)

Decreased *power* to detect RR ≠ 1.0

Can RR < 1.0?

Yes when exposure to an agent reduces the incidence of the illness
E.g., in successful drug trials

Public Policy Reponses to RR estimates are tempered by another consideration.

Given that a factor greatly increases the risk for a disease, how many in the population have that factor?

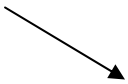
How much of the disease burden in the population can be attributed to that factor?

Attributable Risk, AR in the population =

$$\frac{p(r-1)}{p(r-1)+1}$$

where r = RR, p = proportion of the population with the risk factor (i.e., exposure)

AR tells us that a relative risk of 2 could be very significant if a large fraction of the population has that risk factor



Cancer rates and drinking water, air quality. How much of total cancer burden is due to factors of small RR that effect large populations?
Similarly- diet and heart disease

In contrast- If few people smoked, the RR 10-20 for cigarette smoking would still contribute a lot to overall lung cancer rates. As it is, if smoking prevalence were even 10%, then 50% of the lung cancers would be attributed to smoking.

A final consideration for RR:

What do you know if I tell you that the left half of the class has a RR = 10^6 for receiving a billion $ today compared to the right half of the class? That the left half of the room is associated with greater likelihood of getting money. What do you need to know to evaluate this information?
What do you need to know before you quit school?
What if the right half's "absolute risk" is 10^{-12}?
What if it is 10^{-2}?

We really need to have a measure of absolute risk. More to come on this.