Why Cancer?

Cancer by the Numbers

- Second leading cause of death in the US
  - ~600k deaths/year
  - Leading cause of death in people under age 85
- Less than 1% of US population diagnosed each year
  - ~2 million new cases expected in 2025
- 1/3 of men and women will receive a cancer diagnosis during their lifetime
- ~18 million are living with cancer in the US
- 5-year *relative* survival 69% (all cancers)

https://www.cancer.gov/about-cancer/understanding/statistics
Siegel, et al. Cancer statistics, 2025. CA Cancer J Clin. 2025;75:10–45.

Cancer Biology

- Dysregulated cell growth is the only common thread
- The cancer “phenotype” varies across tissues/organs
  - Presenting symptoms
  - Biological consequences
  - Prognosis
  - Efficacy of treatment
- Diverse molecular biology
- Causes are endogenous and exogenous
- Aging vs. childhood cancer

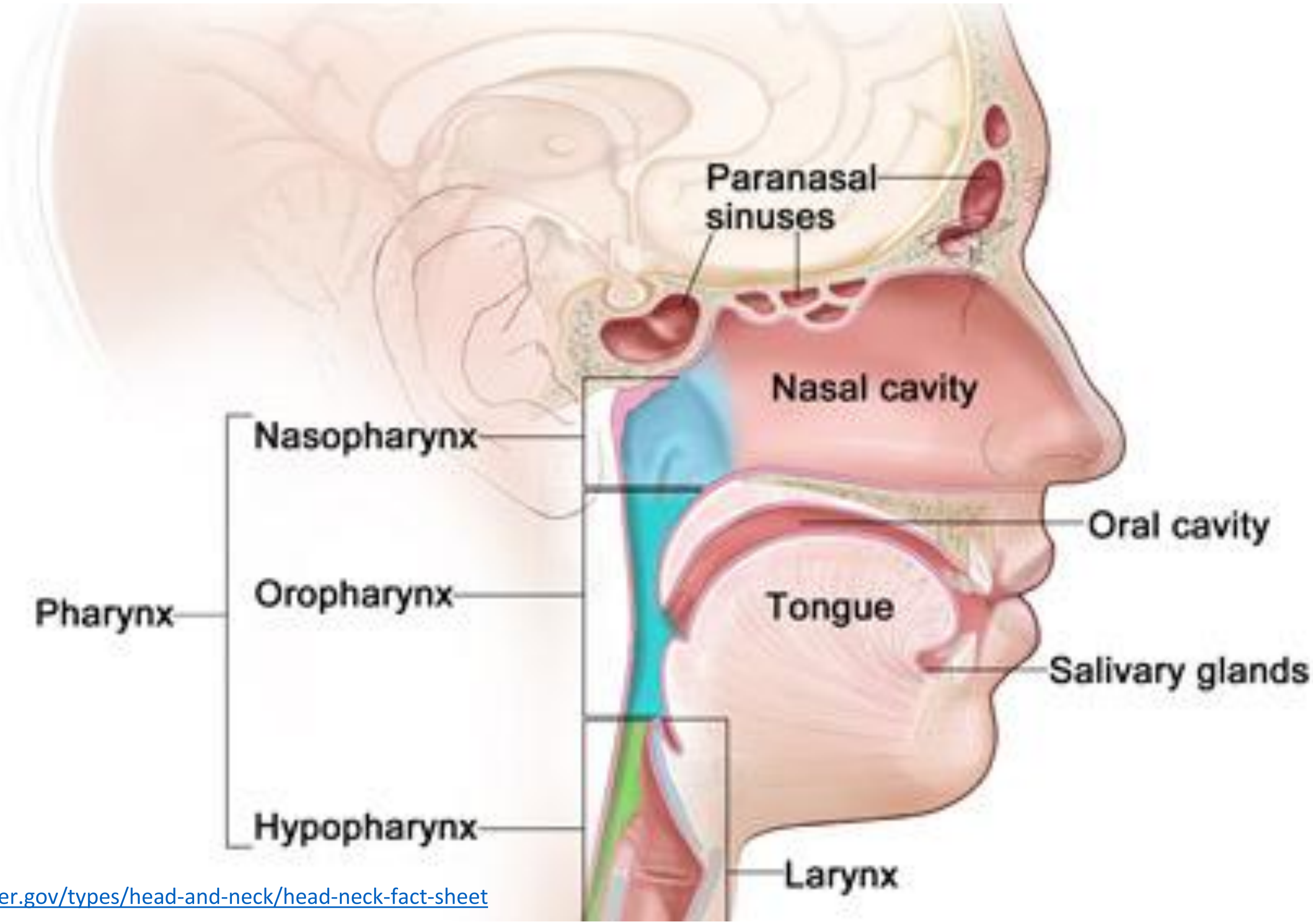
Why Learn from Head and Neck Cancer?

- Known etiology
  - Tobacco, alcohol, HPV infection
  - Gene/environment interactions
- Available treatment
  - Radiation, chemotherapy, biologics
- Well-researched molecular biology
- Excellent examples of all major study designs exist in the literature

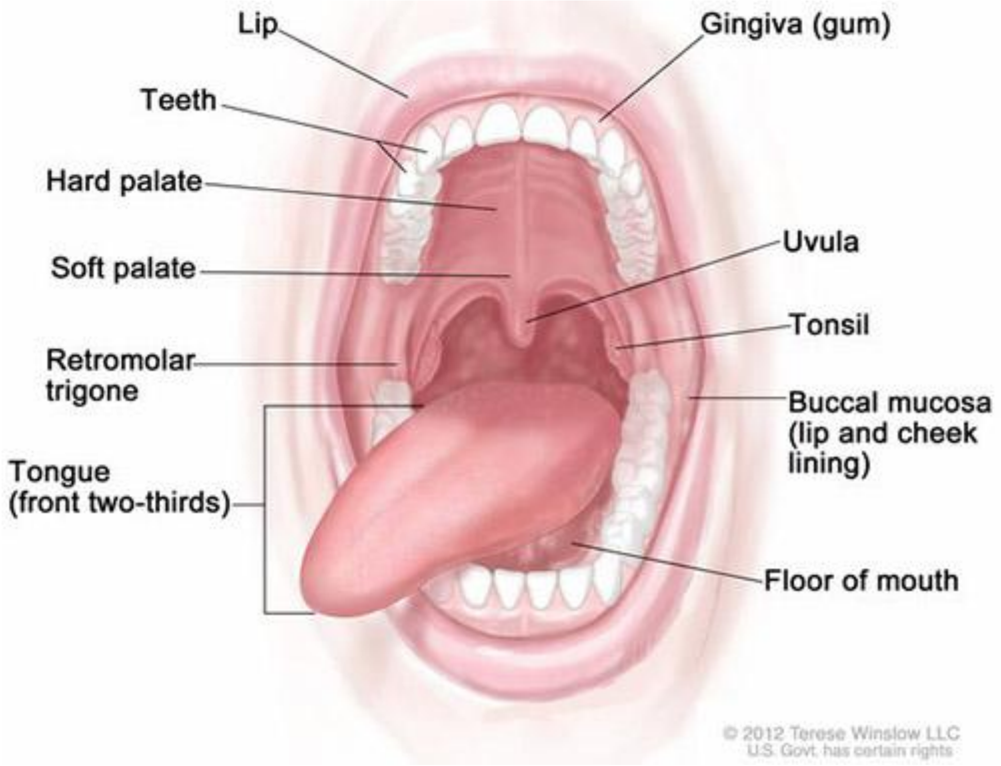
Overview of HNSCC

Head and Neck **S**quamous **C**ell **C**arcinoma

(also referred to as SCCHN)



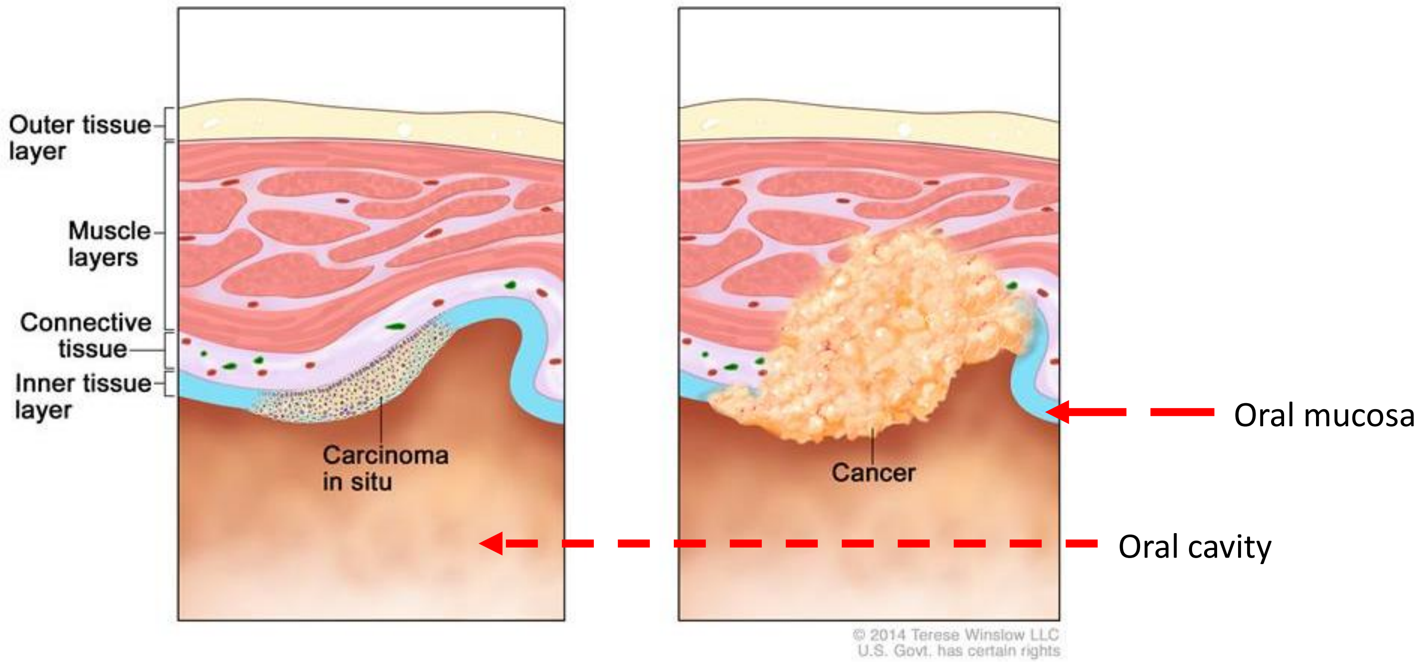
https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet



https://seer.cancer.gov/statfacts/html/oralcav.html

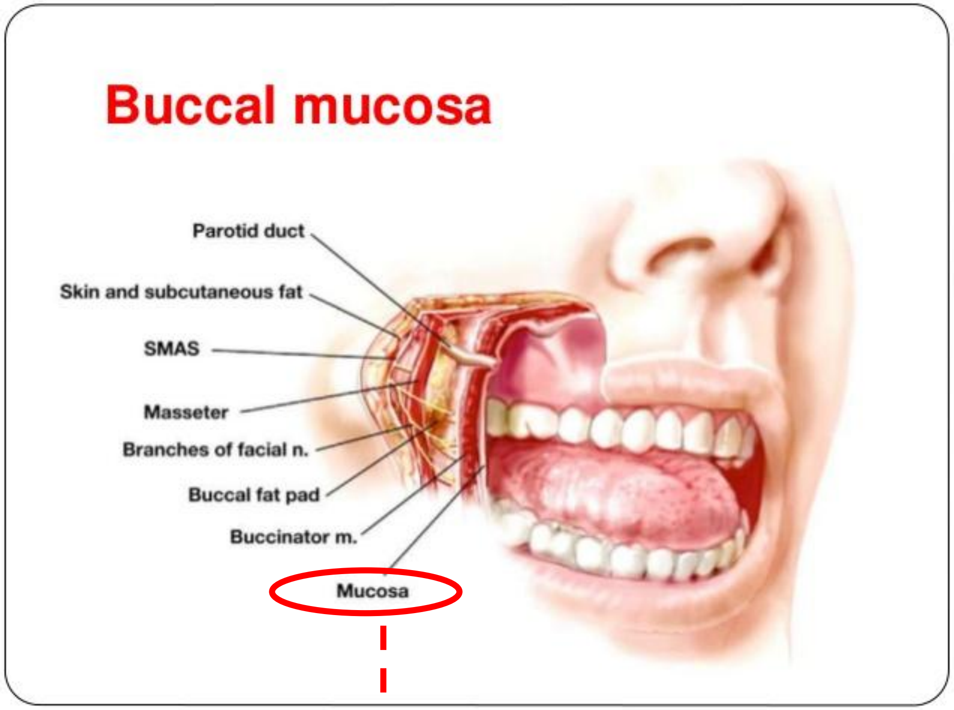
Carcinoma – "cancer that begins in the skin or in tissues that line or cover internal organs"

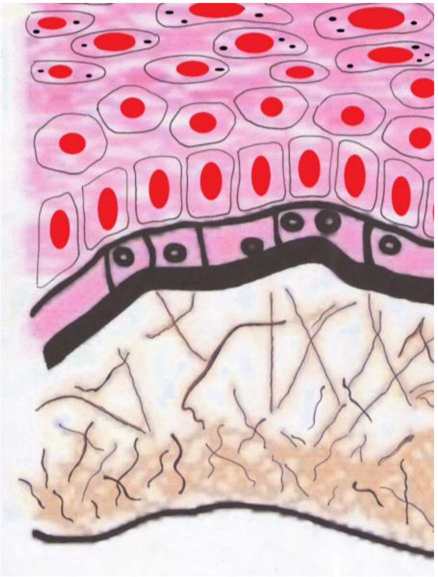
Mucosa – "The moist, inner lining of some organs and body cavities (such as the nose, mouth, lungs, and stomach). Glands in the mucosa make mucus (a thick, slippery fluid). Also called mucous membrane."



https://www.cancer.gov/publications/dictionaries/cancer-terms

Histology
(the microscopic study of tissues)
# of Oral Mucosa



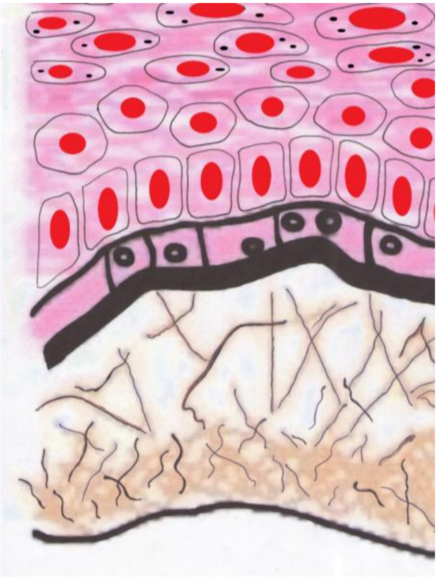


https://image.slidesharecdn.com/carcinomatongue-150622043025-lva1-app6891/95/carcinoma-tongue-4-638.jpg?cb=1434947502

Squamous Cell Carcinoma

Cancer that begins in squamous cells. Squamous cells are thin, flat cells that look like fish scales, and are found in the tissue that forms the surface of the skin, the lining of the hollow organs of the body, and the lining of the respiratory and digestive tracts. Most cancers of the anus, cervix, head and neck, and vagina are squamous cell carcinomas. Also called epidermoid carcinoma.

https://www.cancer.gov/publications/dictionaries/cancer-terms



Squamous epithelium

By Wiki-minor - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=17201584

Adenocarcinomas

- Some mucosal tissues contain *glandular* cells
  - Secrete digestive juices, saliva, mucus, etc.
- Cancer in glandular cells is *adenocarcinoma*
- Head and neck
  - Salivary glands, sinuses
- Other sites where adenocarcinoma is common
  - Breast, lung, colon, prostate

Nasopharyngeal Carcinoma

- Typically squamous cell
- But not included in most studies of HNSCC
- Etiologically distinct
  - Epstein-Barr virus
  - Wood dust
  - Diet heavy in salt-cured fish
- Less associated with alcohol and tobacco
- Geographically distinct incidence (China, Southeast Asia)

https://www.webmd.com/cancer/nasopharyngeal-cancer#1-2

HNSCC Etiology

- Tobacco and alcohol
  - Oral cavity, larynx
- Human papillomavirus (HPV) infection
  - Oropharynx, base of tongue, tonsil
  - ~70% of these in the US are caused by HPV
  - The proportion varies across the world

Nat Rev Microbiol 2012. 10(10):681-692.

HNSCC Epidemiology

Epidemiology:

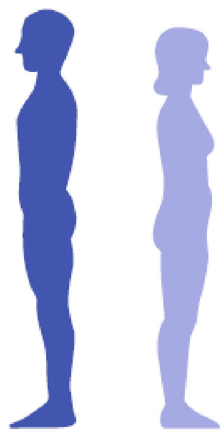
"the branch of medicine which deals with the **incidence**, **distribution**, and **possible control** of diseases and other factors relating to health."

--Oxford English Dictionary

Estimated New Cases, 2024

Estimated New Cases

Male			Female		
Prostate	313,780	30%	Breast	316,950	32%
Lung & bronchus	110,680	11%	Lung & bronchus	115,970	12%
Colon & rectum	82,460	8%	Colon & rectum	71,810	7%
Urinary bladder	65,080	6%	Uterine corpus	69,120	7%
Melanoma of the skin	60,550	6%	Melanoma of the skin	44,410	4%
Kidney & renal pelvis	52,410	5%	Non-Hodgkin lymphoma	35,210	4%
Non-Hodgkin lymphoma	45,140	4%	Pancreas	32,490	3%
Oral cavity & pharynx	42,500	4%	Thyroid	31,350	3%
Leukemia	38,720	4%	Kidney & renal pelvis	28,570	3%
Pancreas	34,950	3%	Leukemia	28,170	3%
All sites	**1,053,250**		**All sites**	**988,660**	



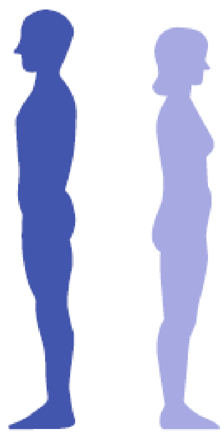
Oral cavity & Pharynx:
Men: 8th most common cancer
Women: Not in the top ten

Siegel, et al. Cancer statistics, 2025. CA Cancer J Clin. 2025;75:10–45.

Estimated Cancer Deaths, 2022

Estimated Deaths

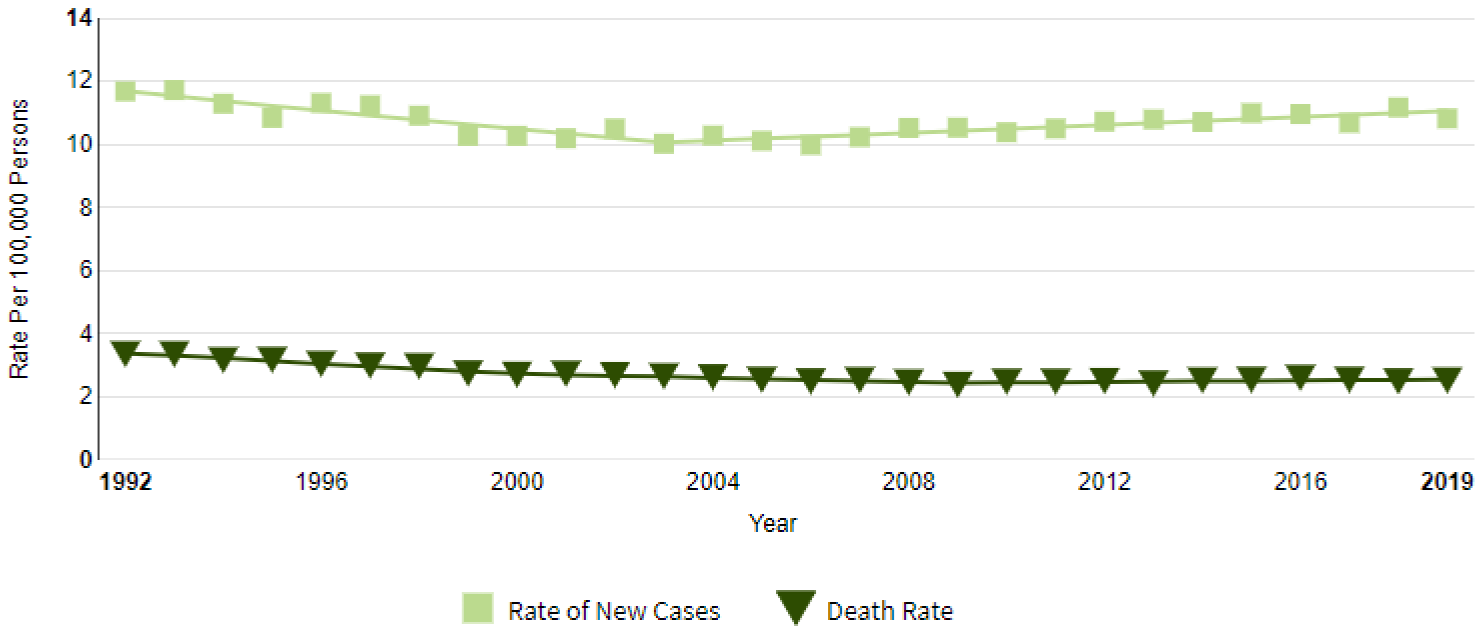
Male				Female		
Lung & bronchus	64,190	20%		Lung & bronchus	60,540	21%
Prostate	35,770	11%		Breast	42,170	14%
Colon & rectum	28,900	9%		Pancreas	24,930	8%
Pancreas	27,050	8%		Colon & rectum	24,000	8%
Liver & intrahepatic bile duct	19,250	6%		Uterine corpus	13,860	5%
Leukemia	13,500	4%		Ovary	12,730	4%
Esophagus	12,940	4%		Liver & intrahepatic bile duct	10,840	4%
Urinary bladder	12,640	4%		Leukemia	10,040	3%
Non-Hodgkin lymphoma	11,060	3%		Non-Hodgkin lymphoma	8,330	3%
Brain & other nervous system	10,170	3%		Brain & other nervous system	8,160	3%
All sites	**323,900**			**All sites**	**294,220**	



Oral cavity & Pharynx cancers are not a major cause of cancer deaths for men or women

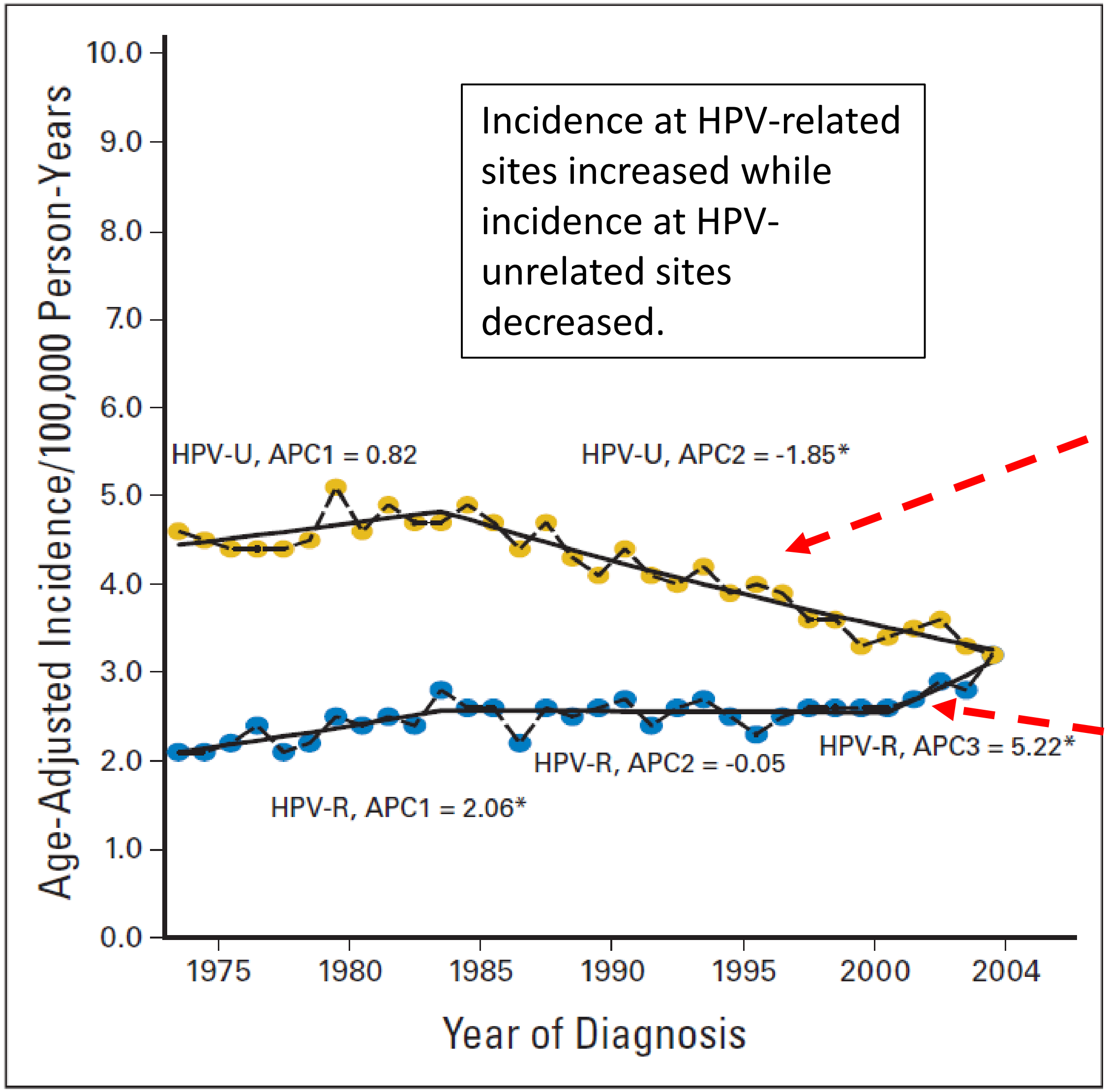
Siegel, et al. Cancer statistics, 2025. CA Cancer J Clin. 2025;75:10–45.

Oral Cavity & Pharynx



https://seer.cancer.gov/statfacts/html/oralcav.html

Incidence Trends: Oral & Pharyngeal



Anterior tongue, oral cavity

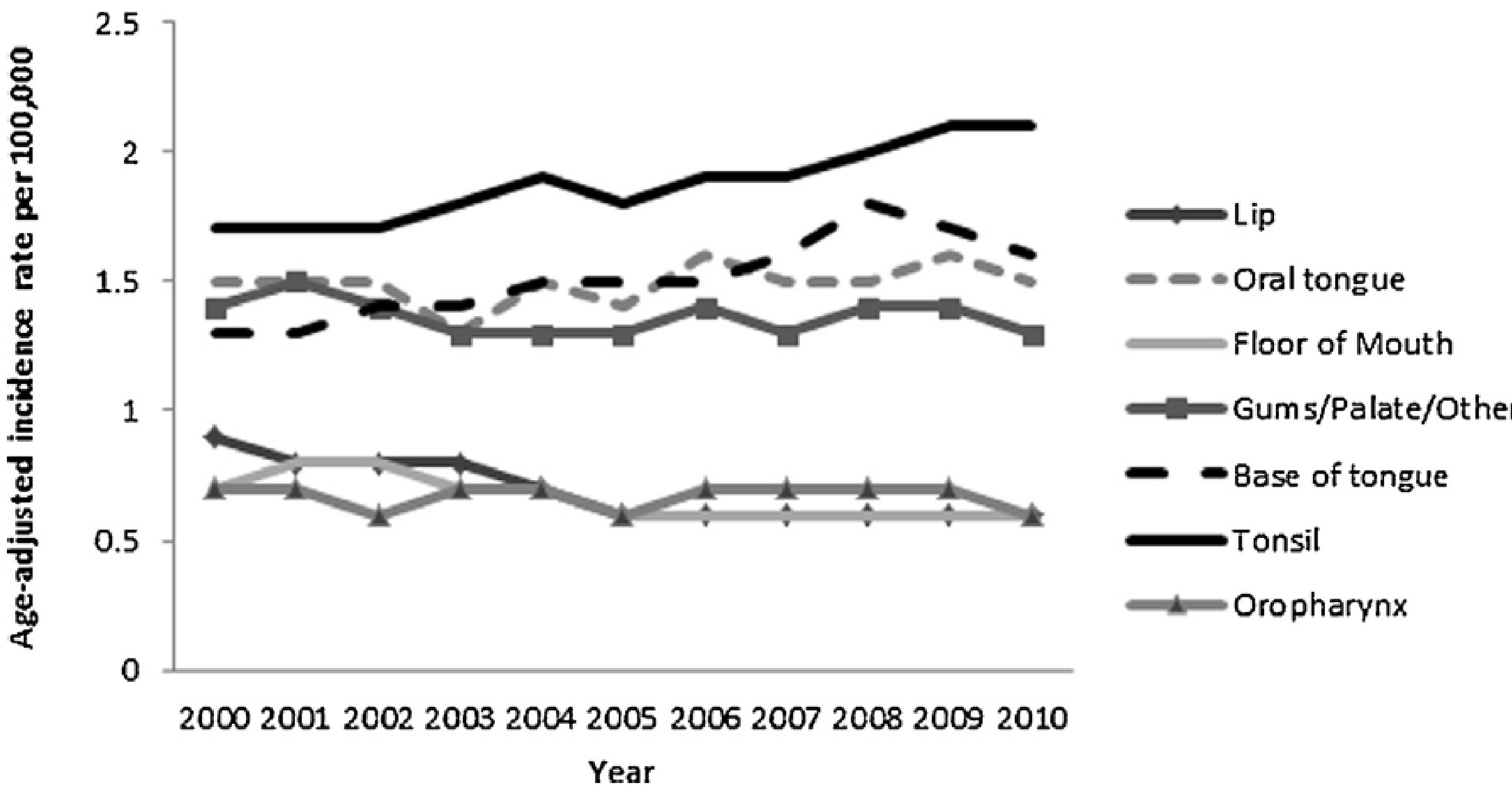
Oropharynx, tonsil, base of tongue

J Clin Oncol 26:612-619.

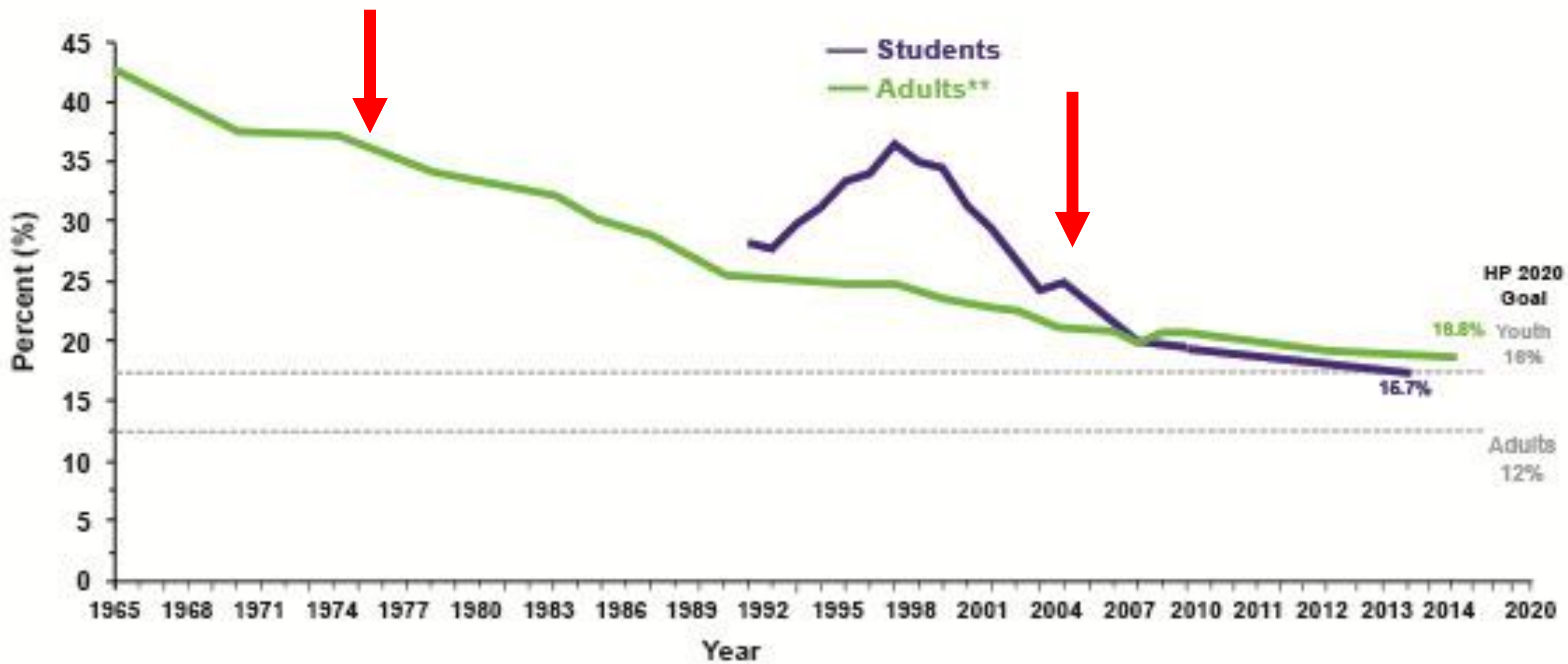
Incidence Trends by Site

Increases in HPV-related sites were mostly in the tonsil and base of tongue.

D.J. Weatherspoon et al./Cancer Epidemiology 39 (2015) 497–504



Trends in Current Cigarette Smoking by High School Students* and Adults** — United States, 1965-2014



*Percentage of high school students who smoked cigarettes on 1 or more of the 30 days preceding the survey (Youth Risk Behavior Survey, 1991-2013).
**Percentage of adults who are current cigarette smokers (National Health Interview Survey, 1965-2014).

https://www.cdc.gov/tobacco/data_statistics/tables/trends/cig_smoking/index.htm

Information Gained

- Trends in the incidence of a disease can suggest a shift in distribution of the causal agent in the population
- How then might we define a *cause*?

"A factor is a cause of an event if its operation increases [or decreases] the frequency of the event"

--Elwood

Level of Measurement

- Ecologic
  - Exposure (smoking) and outcome (cancer) are measured at the same time in the same population, but not necessarily in the same people
- Individual Level
  - Observation of decreased incidence of HNSCC in individuals who don’t smoke vs. those who do smoke
  - Exposure (smoking) and outcome (cancer) are measured in the same people

Measuring Frequency: Prevalence

- Number of people living with HNSCC during a specific time period divided by the size of the population at the beginning of the period
- 100 people were living with HNSCC in Durham, NC during 2016-2018
- The population of Durham was 263,000 in 2016
- Prevalence = 100/263,000 = 0.04%

Measuring Frequency: Incidence

- Identifies only the newly occurring cases during a specific time period
- Incidence proportion (cumulative incidence)

$$\frac{\#\ of\ new\ cases\ diagnosed\ from\ time\ A\ to\ B}{size\ of\ population\ at\ risk\ at\ time\ A}$$

There were 50 incident cases of HNSCC in Durham during 2016-2018 among 262,900 at risk at the beginning of the year (0.02%)

Incidence Rate (approximation)

$$\frac{\#\ of\ new\ cases\ diagnosed\ from\ time\ A\ to\ B}{average\ population\ size\ during\ time\ A\ to\ B}$$

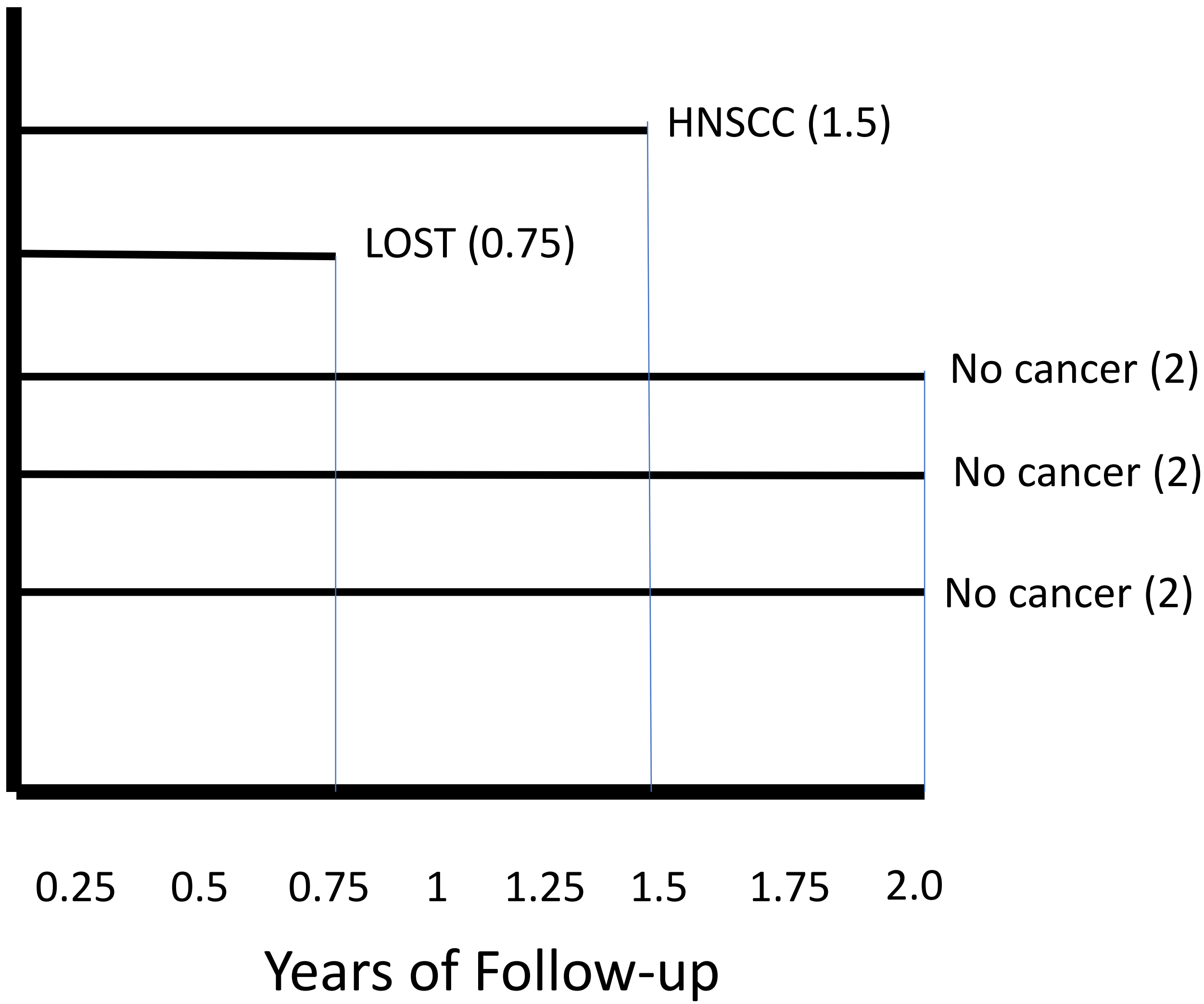
There were 50 new cases of HNSCC diagnosed in Durham, NC during 2016-2018 among an average population of 262,598 people

2016: 263,000 people

2018: 262,195 people

(50/262,598)*100,000 ~= 19 cases per 100,000 persons/year

Incidence Rate (exact)



Incidence Rate:

1 case / 8.25 person-years
or:

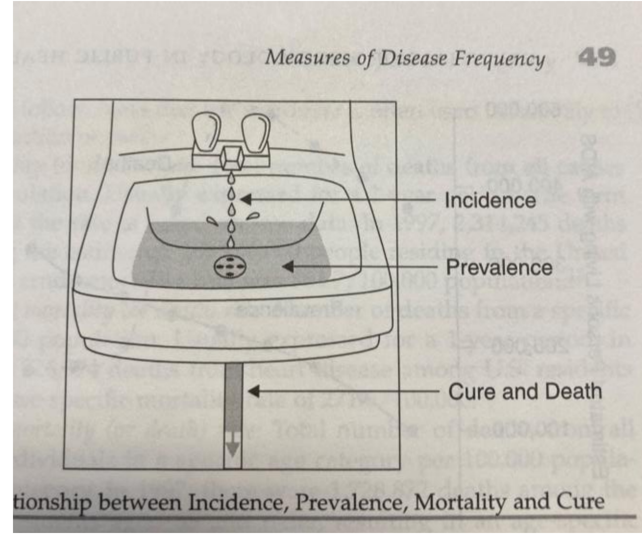
(1/8.25)*1,000 =

~121 cases/1,000 person-years

or

121 cases per 1,000 persons per year

P=ID



Previously we used the example:

$I_{2016\text{-}2018} = 50/262{,}900$
$P_{2016\text{-}2018} = 100/263{,}000$

Average duration (time surviving with HNSCC) is:

$$D = \frac{P}{I} = \frac{(100/263{,}000)}{(50/262{,}900)} = 2 \text{ years}$$

This makes a lot of assumptions but is illustrative of the basic relationships between prevalence and incidence and duration.

Image is from *Essentials of Epidemiology in Public Health* by Aschengrau and Seage. 2003. Page 49.

Incidence is Preferred

- Because P=ID
  - Prevalence includes information about new cases (I)
  - And survival after diagnosis (D)
- Studies that use P as a measure of frequency
  - Evaluate causes of disease (I); and
  - Evaluate determinants of survival (D)
- The preferred measure of frequency for studying *cause* is incidence (I)
- Best calculated as an exact rate per unit of person-time when possible

Study Design

"Does cigarette smoking cause HNSCC?"

A General Approach

- Identify some smokers
  - How long? How much?
- Identify some non-smokers
  - Never? Or a little?
- Compare the incidence of HNSCC in these groups
  - Over what period of time?
- Smokers should have a higher incidence of HNSCC
  - How much higher?
- Are there any other reasons why smokers have a higher incidence of HNSCC?
  - Are they older?

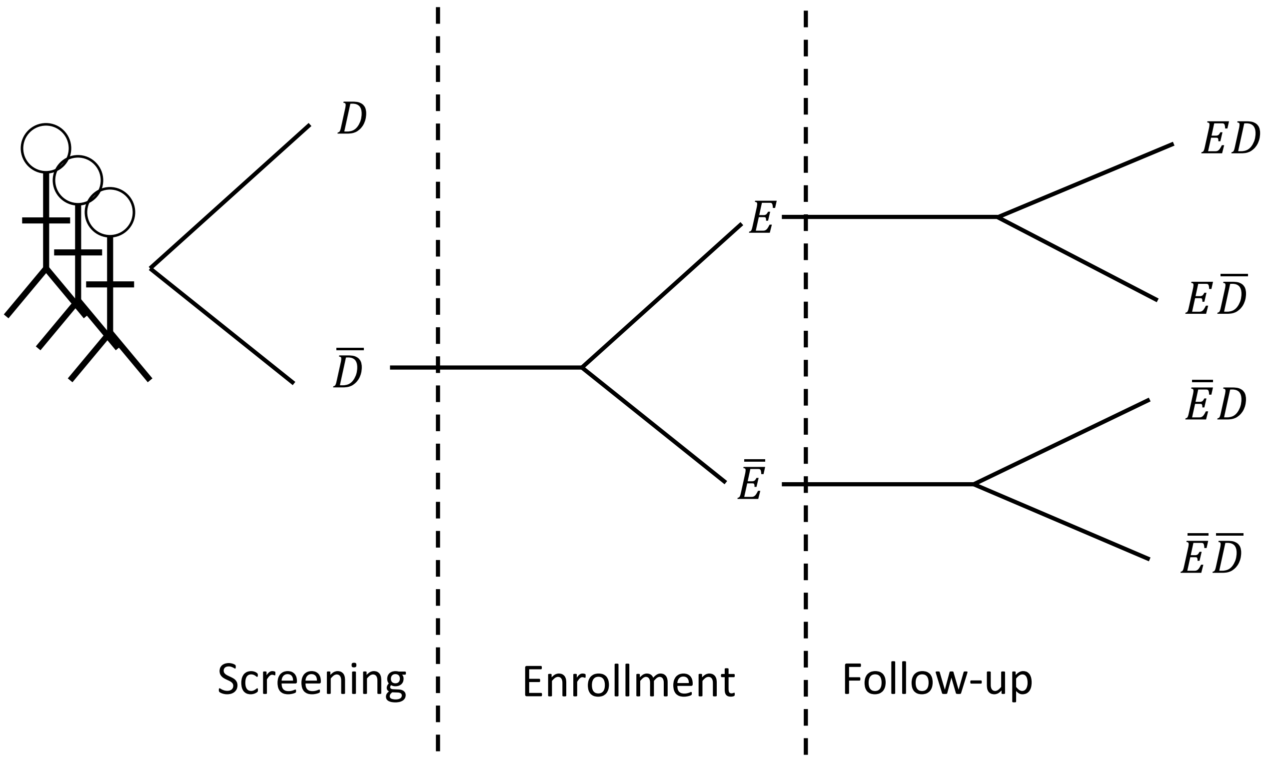
Salient Features of this Approach

- Based on a quantifiable definition of *cause*
  - A cause *increases the frequency* of an event
  - **Risk**; not determinism, sufficiency or necessity
- Uses a comparison group
  - The frequency (incidence) of HNSCC in smokers is compared to the same rate in non-smokers
- Uses a *measure of association* to express the relationship between smokers and non-smokers
  - Incidence is “higher” in smokers than non-smokers
- Considers possible sources of mistaken conclusion
  - Cancer is more common in older people
  - If smokers are older than non-smokers then the incidence of HNSCC may appear “higher” due to difference in age distribution

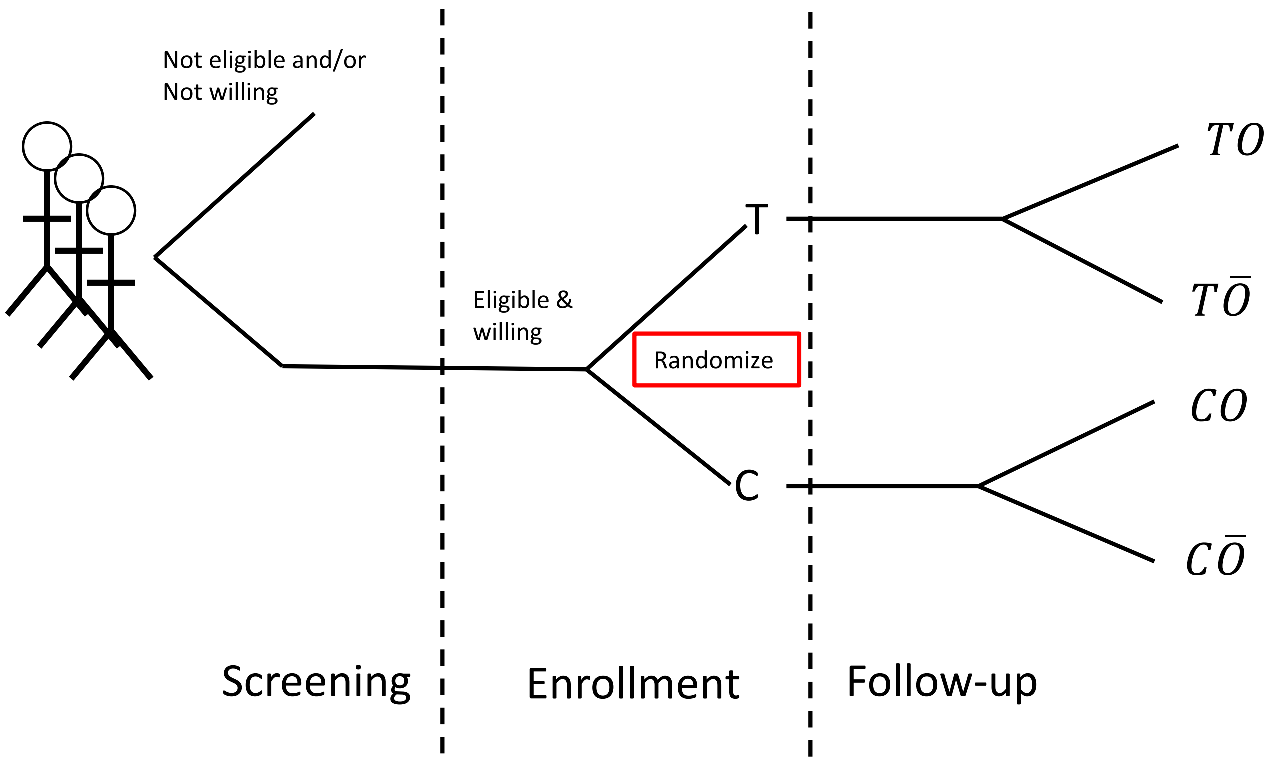
Study Design

- Study designs provide a framework for causal inference that facilitates
  - Quantitative analysis
  - Meaningful interpretation
  - Guarding against mistaken conclusion
- Each design has strengths/weaknesses in specific contexts
- All designs can yield valid and useful results

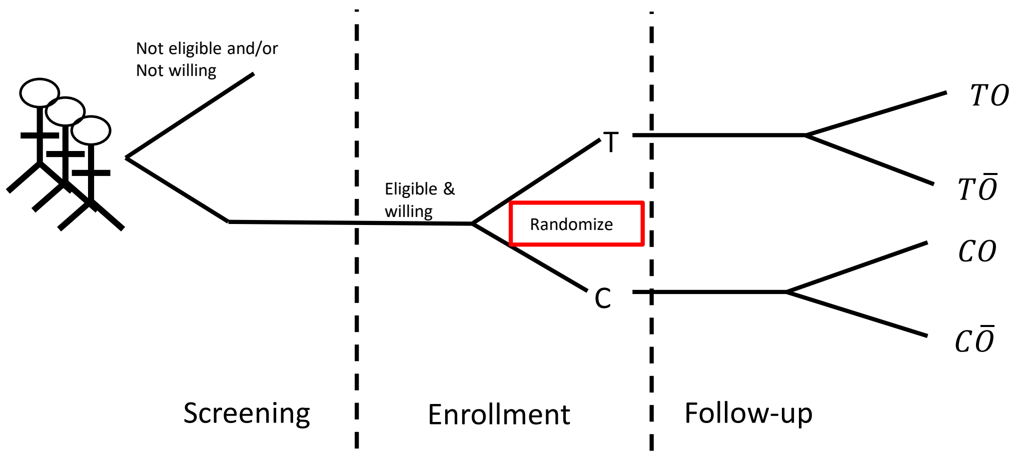
Cohort Study Design



Clinical Trial Design



Measures of Association



Cum. Incid. of O on treatment = $\dfrac{TO}{TO+T\bar{O}}$ = risk of O in treated patients = r_t

Cum. Incid. of O on control = $\dfrac{CO}{CO+C\bar{O}}$ = risk of O in control patients = r_c

Risk difference (RD) = $r_t - r_c$

Relative risk (RR) = r_t / r_c

Interpretation of the RD & RR

- RD = $r_t - r_c$
  - < 0 risk of outcome is lower on treatment than control
  - 0 no difference in risk of outcome comparing *t* vs. *c*
  - > 0 risk of outcome is higher on *t* than *c*
- Relative risk (RR) = r_t / r_c
  - < 1 risk of outcome is lower on *t* than *c*
  - 1 no difference in risk of outcome comparing *t* vs. *c*
  - > 1 risk of outcome is higher on *t* than *c*
  - The RR is never negative (and is seldom equal to zero)

Example - Calculation

Risk of second primary malignancy after successful treatment of HNSCC

	Second Primary Malignancy	No Second Primary Malignancy	*Total (N=841)*
Male	43	468	511
Female	17	313	330

$$Risk_{males} = 43/511 = 0.0841$$

$$Risk_{female} = 17/330 = 0.0515$$

$$RD_{male-female} = 0.0841 - 0.0515 = 0.0326$$

$$RR_{male\ vs.female} = 0.0841/0.0515 = 1.63$$

Example - Interpretation

$$Risk_{males} = 43/511 = 0.0841$$

$$Risk_{female} = 17/330 = 0.0515$$

The risk of second primary malignancy after successful treatment for HNSCC is 8.41% in men and 5.15% in women.

$$RD_{male-female} = 0.0841 - 0.0515 = 0.0326$$

The risk of second primary malignancy after successful treatment for HNSCC is is 3.26% higher in men than women.

$$RR_{male\ vs.female} = 0.0841/0.0515 = 1.63$$

The risk of second primary malignancy after successful treatment for HNSCC is 1.63 times as high in men as it is in women.

The risk of second primary malignancy after successful treatment for HNSCC is 63% higher in men than women.

RD, RR and Prevalence

	Second Primary Malignancy	No Second Primary Malignancy	*Total (N=841)*
Male	43	468	511
Female	17	313	330

$$Prevalence = \frac{43 + 17}{511 + 330} = 0.07$$

RD=0.03, RR=1.63

	Second Primary Malignancy	No Second Primary Malignancy	*Total (N=400)*
Male	35	165	200
Female	22	178	200

$$Prevalence = \frac{35+22}{200+200} = 0.14$$

RD=0.07, RR=1.60

Be Careful!

	Second Primary Malignancy (+)	No Second Primary Malignancy (-)	*Total*
Male	43	468	511
Female	17	313	330

$$\boxed{\frac{1}{RR_{(+)\ a\ vs.\ b}} = RR_{(+)\ b\ vs.\ a}}$$

$$RR_{(+)\ male\ vs. female} = \frac{\left({}^{43}/_{511}\right)}{\left({}^{17}/_{330}\right)} = \frac{(43)(330)}{(511)(17)}$$

$$\frac{1}{RR_{(+)\ male\ vs. female}} = \frac{\left({}^{17}/_{330}\right)}{\left({}^{43}/_{511}\right)} = \frac{(511)(17)}{(43)(330)} = RR_{(+)\ female\ vs. male}$$

$$\boxed{\frac{1}{RR_{(+)\ a\ vs.\ b}} \neq RR_{(-)\ a\ vs.\ b}}$$

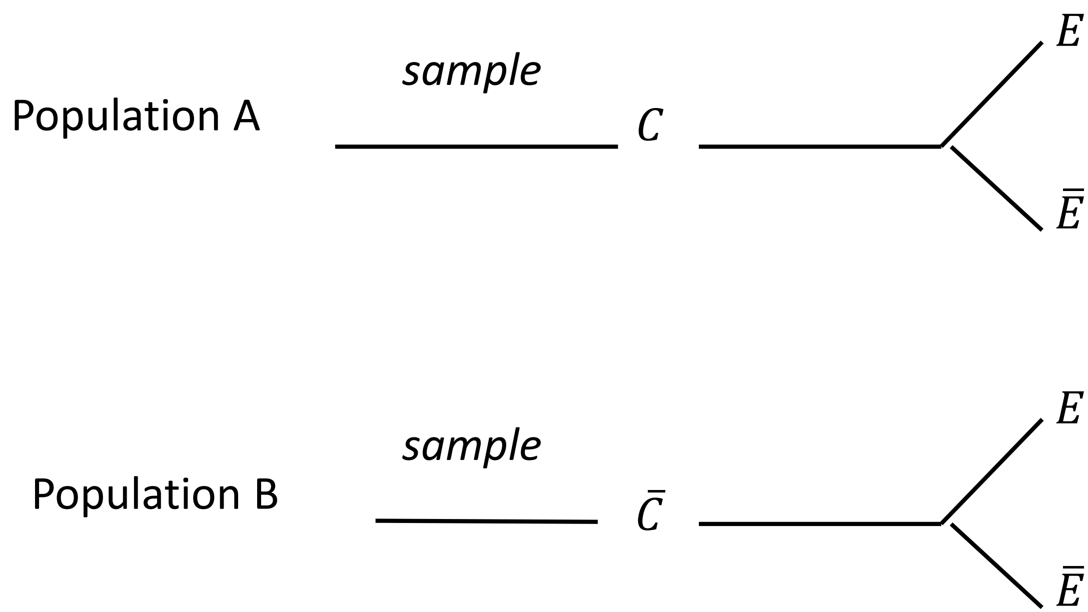
$$RR_{(+)\ male\ vs. female} = \frac{(43)(330)}{(511)(17)} = 1.63 \text{ and } \frac{1}{1.63} = 0.613$$

$$RR_{(-)\ male\ vs. female} = \frac{{}^{468}/_{511}}{{}^{313}/_{330}} = \frac{(468)(330)}{(511)(313)} = 0.96 \neq 0.613$$

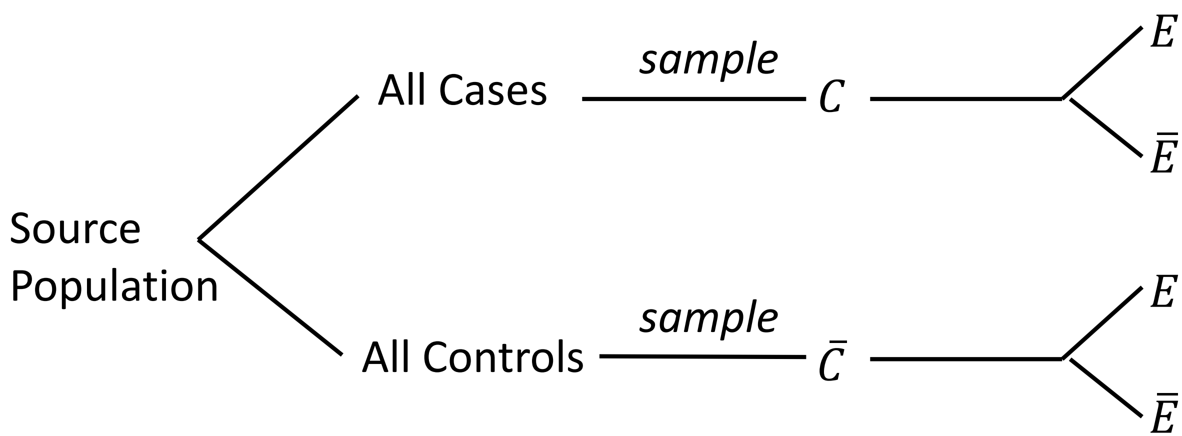
If you invert the RR of the outcome for groups a vs. b then you get the RR of the outcome for groups b vs. a.

Inverting the RR does *not* give you the RR of <u>not</u> having the outcome!

Case-Control Design

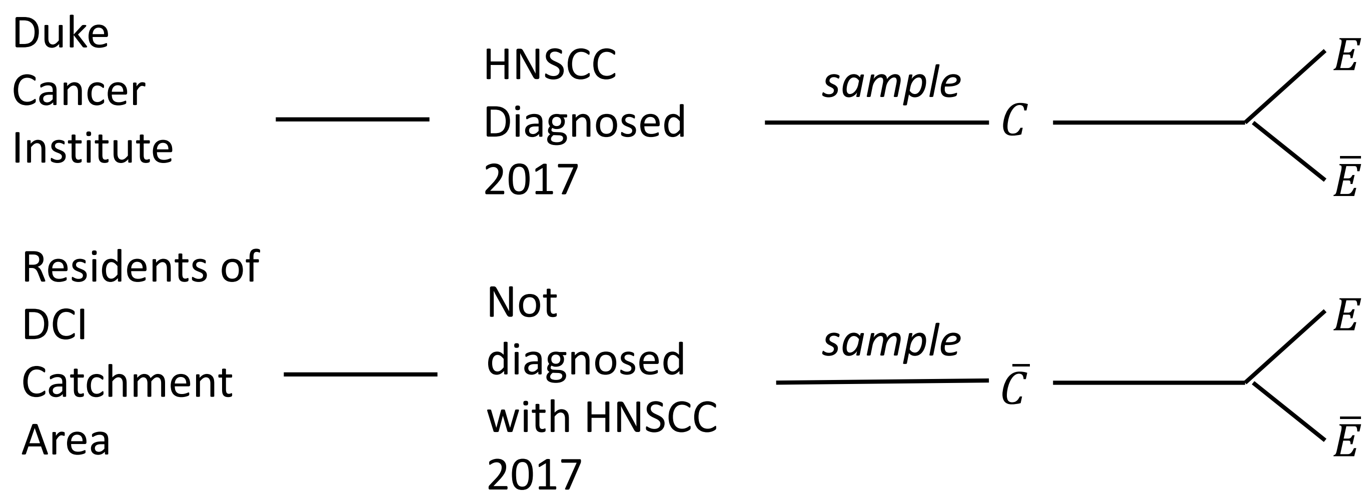


Population Based Case-Control



Example

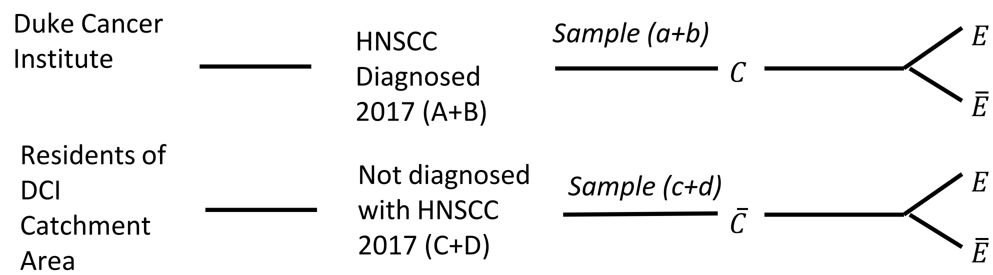
Does cigarette smoking cause HNSCC?



	Smokers (E)	Non-Smokers $\bar{E}$
HNSCC (C)	a	b
No HNSCC ($\bar{C}$)	c	d

Problem

Does cigarette smoking cause HNSCC?



	Smokers (E)	Non-Smokers $\bar{E}$
HNSCC (C)	a	b
No HNSCC ($\bar{C}$)	c	d

Estimation of risk using our definition of “cause” implies that you can enumerate the population(s) you start with:

$$RR = \frac{frequency\ of\ exposure\ in\ HNSCC\ Cases}{frequency\ of\ exposure\ in\ Controls} = \frac{a/(a+b)}{c/(c+d)}$$

In a case-control study we *sample* cases and controls and do not know the true size of the denominators that the cases and controls arose from (a+c and b+d). Therefore it is not appropriate to measure “risk” in exposed and unexposed.

Thus, RR is not an appropriate measure of association (RD has a similar problem)

Odds Ratio

	Smoker (exposed)	Non-Smoker (unexposed)	*Total*
HNSCC (cases)	43	35	78
No HNSCC (controls)	17	25	42

$$OR = \frac{43/35}{17/25} = 1.81$$

The odds of smoking is 1.81 times higher HNSCC patients compared with controls

$$OR = \frac{43/17}{35/25} = 1.81$$

The odds of HNSCC is 1.81 times higher in smokers compared with non-smokers

Interpretation of the OR

- Disease odds ratio
  - Odds ratio (OR) = $o_{Exposed}$ / $O_{Unexposed}$
  - < 1 odds of disease is lower in *exposed* than *unexposed*
  - 1 no difference in odds of disease
  - > 1 odds of disease is higher in *exposed* than *unexposed*
- Exposure odds ratio
  - Odds ratio (OR) = o_{Cases} / $O_{Controls}$
  - < 1 odds of exposure is lower in *Cases* than *Controls*
  - 1 no difference in odds of exposure
  - > 1 odds of exposure is higher in *Cases* than *Controls*
- The OR is never negative (and is seldom zero)

Inverting the OR

	Smokers (E)	Non-Smokers $\bar{E}$
HNSCC (C)	a	b
No HNSCC ($\bar{C}$)	c	d
Total	N_1	N_2

$$OR = \frac{disease\ odds\ in\ smokers}{disease\ odds\ in\ non-smokers} = \frac{\left({}^{a}/_{c}\right)}{\left({}^{b}/_{d}\right)}$$

$$\frac{1}{\mathrm{OR}} = \frac{\left({}^{c}/_{a}\right)}{\left({}^{d}/_{b}\right)} = \frac{odds\ of\ \boldsymbol{no\ disease}\ in\ smokers}{odds\ of\ \boldsymbol{no\ disease}\ in\ non-smokers}$$

The odds ratio of an event is the reciprocal of the odds ratio of a non-event

This property is unique to the OR and is not shared by the RR!

OR and Rarity of Disease
(to be covered in detail later)

- The odds approximates probability for rare diseases
- But you can still do a case-control study with a common disease
- The OR is always a valid measure of association
- More on this in the Case-control module!