# Cohort studies

**Hans Wolff**

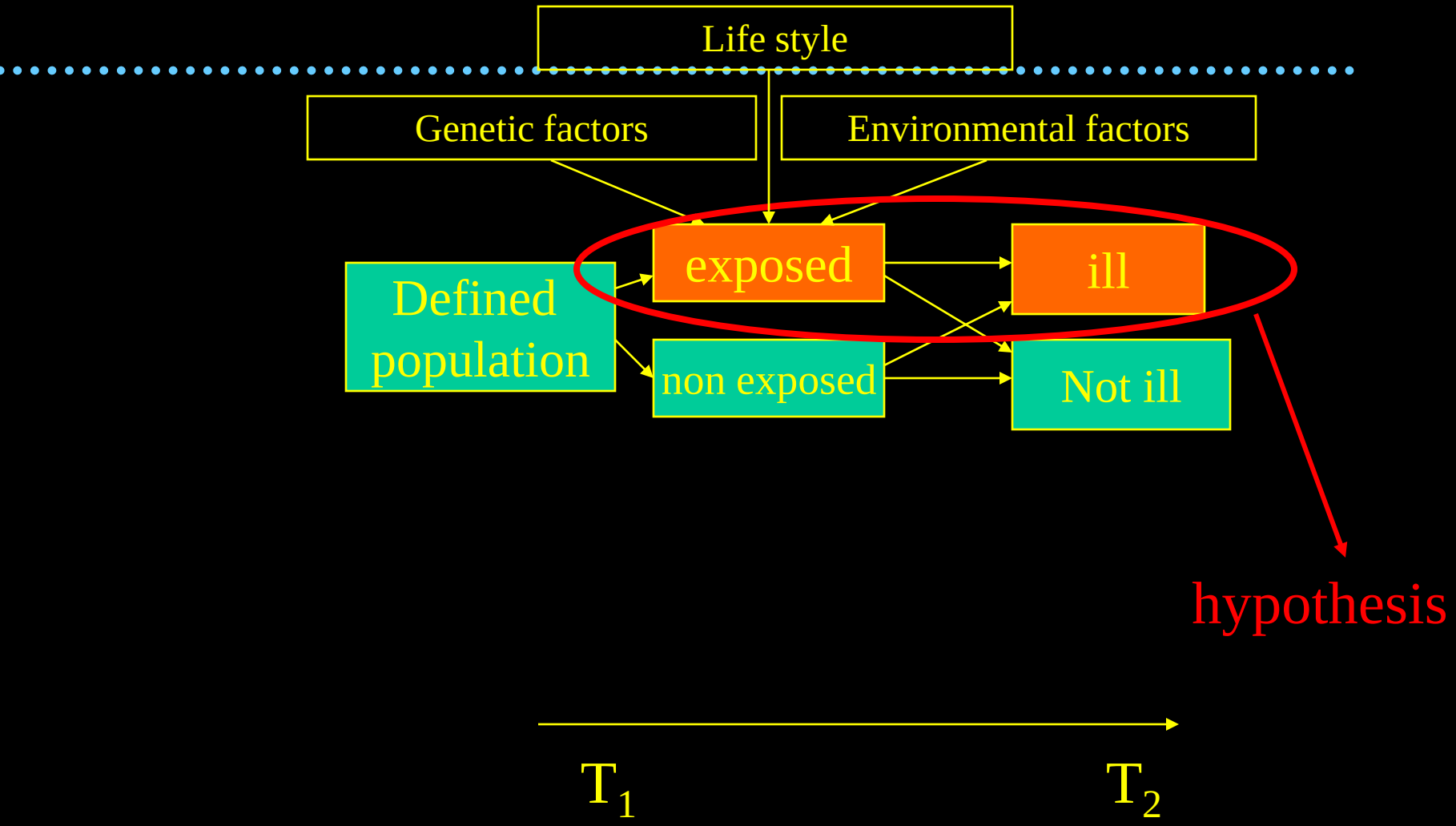
**Unité d'épidémiologie Populationnelle,**

**Département de médecine communautaire et de premier recours**

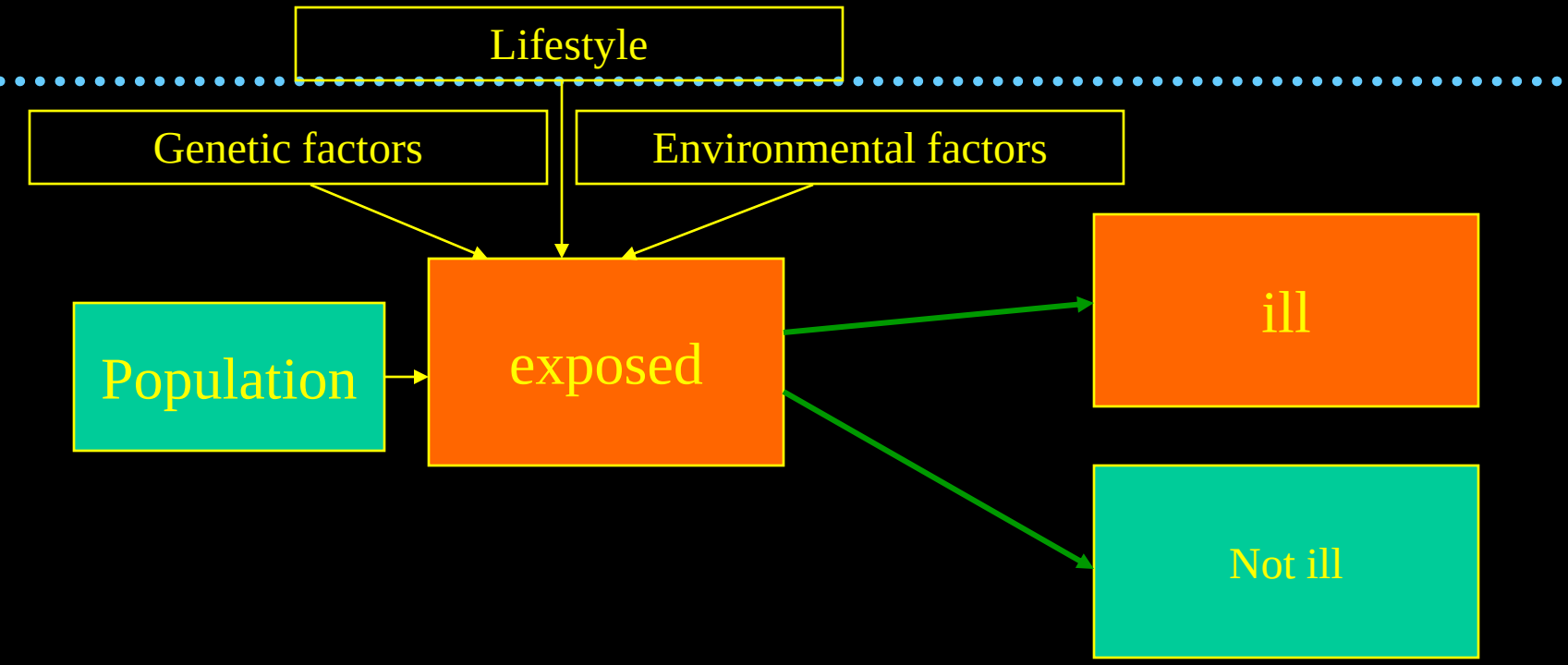
**Hans.Wolff@hcuge.ch**

HUG
Hôpitaux Universitaires de Genève

# Cohort study (CS)



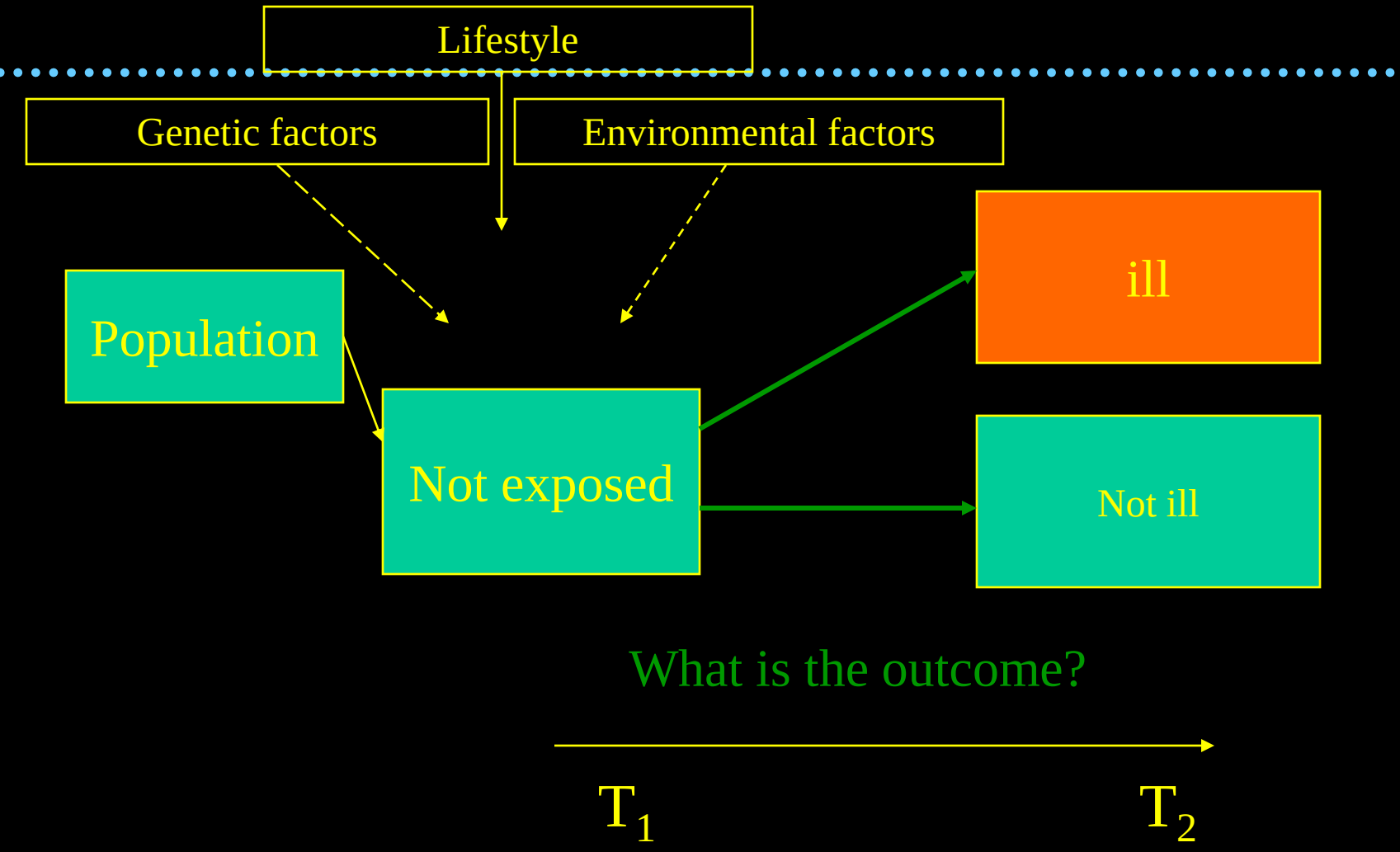
# Cohort study (CS)




What is the outcome ?

$T_1$ $T_2$

# Cohort study (CS)

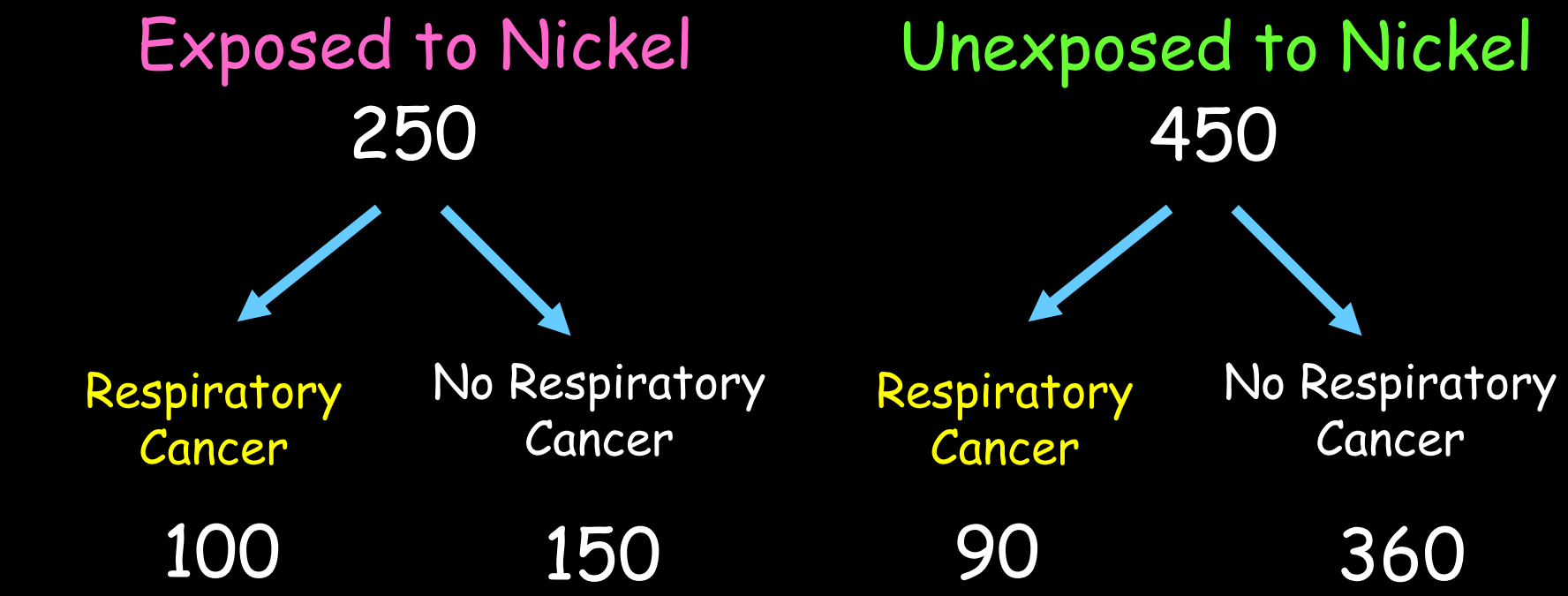


# Outline

- Working Example
  - Welsh Nickel Workers Study
  - Description of the study and raw data in…
    - Breslow, N.E., Day N.E. Statistical Methods in Cancer Research. IARC, 1987:369-74

# Cohort Design

## SOUTH WALES REFINERY WORKERS



# Example

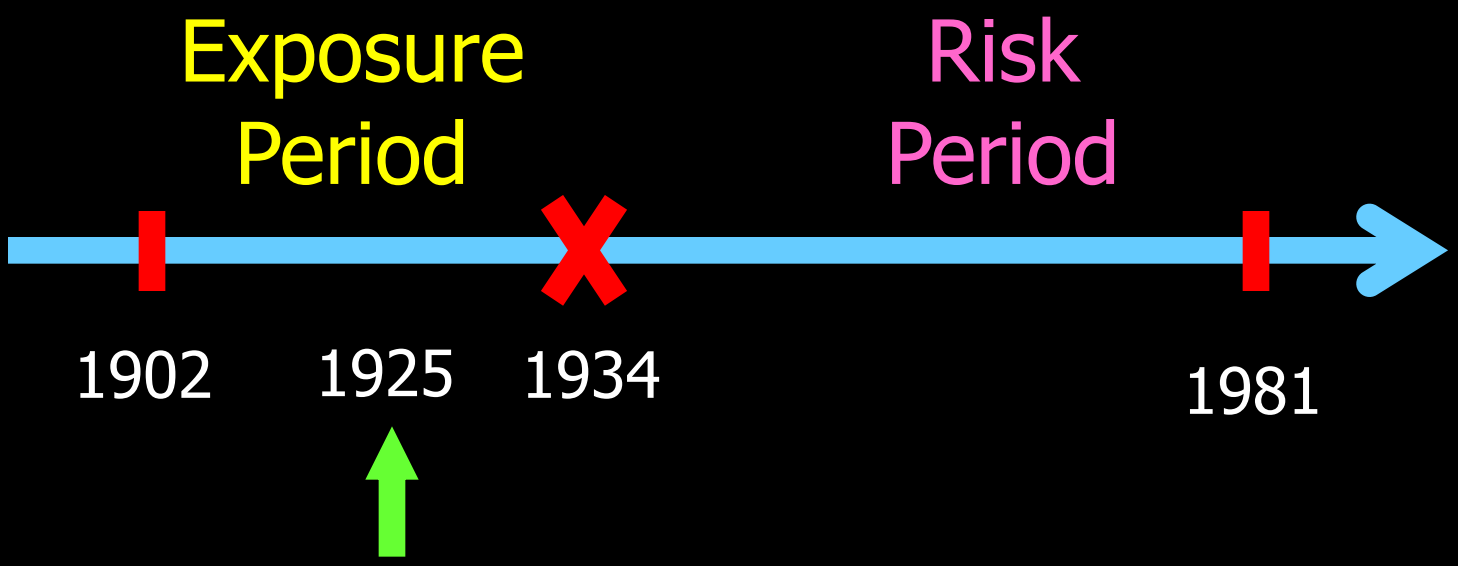
| | 250 Exposed To Nickel | 450 Unexposed To Nickel |
|---|---|---|
| Respiratory Cancer | 100 | 90 |
| Person-years | 4,100 | 11,000 |
| Incidence Rate | 0.024/yr | 0.008/yr |
| Relative Incidence rate | 3.0 | |
| Attributable Risk | 0.016/yr | |

# Study design

- Population: a Nickel factory of South Wales
- Nickel production by decomposition of gaseous nickel compounds
- Exposure: according to information on jobs at high risk of exposure held from 1902 to 1934
- Risk period: count cases of RC* between April 1934 to December 1981
- Outcome: respiratory, mostly lung and nasal cancer

* RC = respiratory cancer

# Study design



# Which is a fundamental condition for the validity of this cohort design ?

- Subjects need to be:
  1. A random sample of the population?
  2. At risk of developing lung or nasal cancer ?
  3. Unlikely to get colon cancer ?
  4. Randomized to nickel exposure ?
  5. Willing to answer questionnaires for many years ?

# "At risk of Respiratory Cancer"

- Never had respiratory cancer: exclude prevalent cases
- Still have two lungs … and a nose: exclude subjects who cannot travel from the denominator to the numerator

# "Incident Respiratory Cancer"

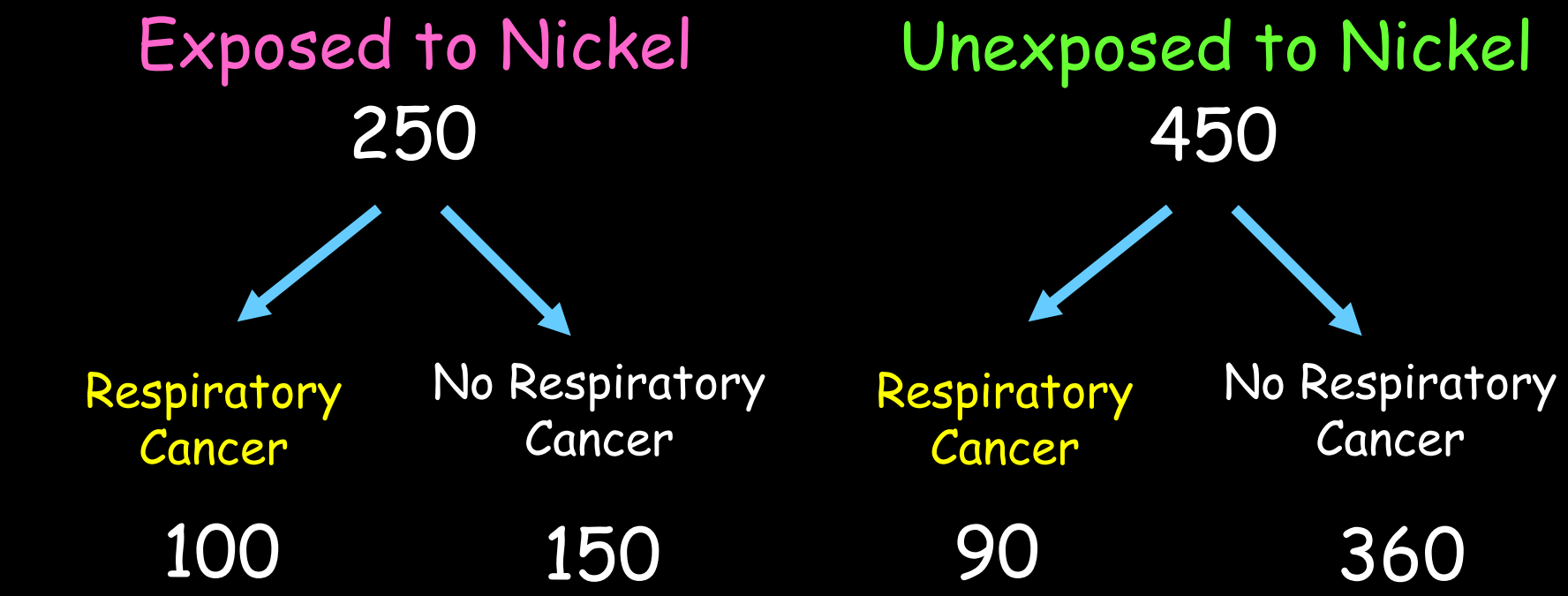
- Incident = "newly diagnosed"
- Between April 1,1934 and December 31,1981
  - Risk Period = 47 years
- Employed in the factory before 1925

# What is the risk of respiratory cancer in this study ?

1. Probability of developing RC per 100,000 workers and per year
2. Probability of developing RC over 47 years
3. The excess probability of RC due to exposure
4. The ratio of the probability of RC in exposed over the probability of RC in unexposed
5. A synonymous for the odds of RC

# Cohort Design

## SOUTH WALES REFINERY WORKERS



# Risk of respiratory cancer in unexposed

| | Unexposed to Nickel |
|---|---|
| Respiratory Cancer | 90 |
| Total | 450 |
| Person-years | 11,000 |

Risk =

**Interpretation:**

# What is the risk of respiratory cancer in unexposed ?

1. $\left(\frac{90}{450}\right)$
2. $\left(\frac{90}{450-90}\right)$
3. $\left(\frac{450-90}{450}\right)$
4. $\left(\frac{90}{11{,}000}\right)$
5. $\left(\frac{90}{11{,}000 - 90}\right)$

# Calculating Risk in Unexposed

$$\text{Risk}_{\text{time}} = \left( \frac{\text{New events}}{\text{Population "at risk" at baseline}} \right)$$

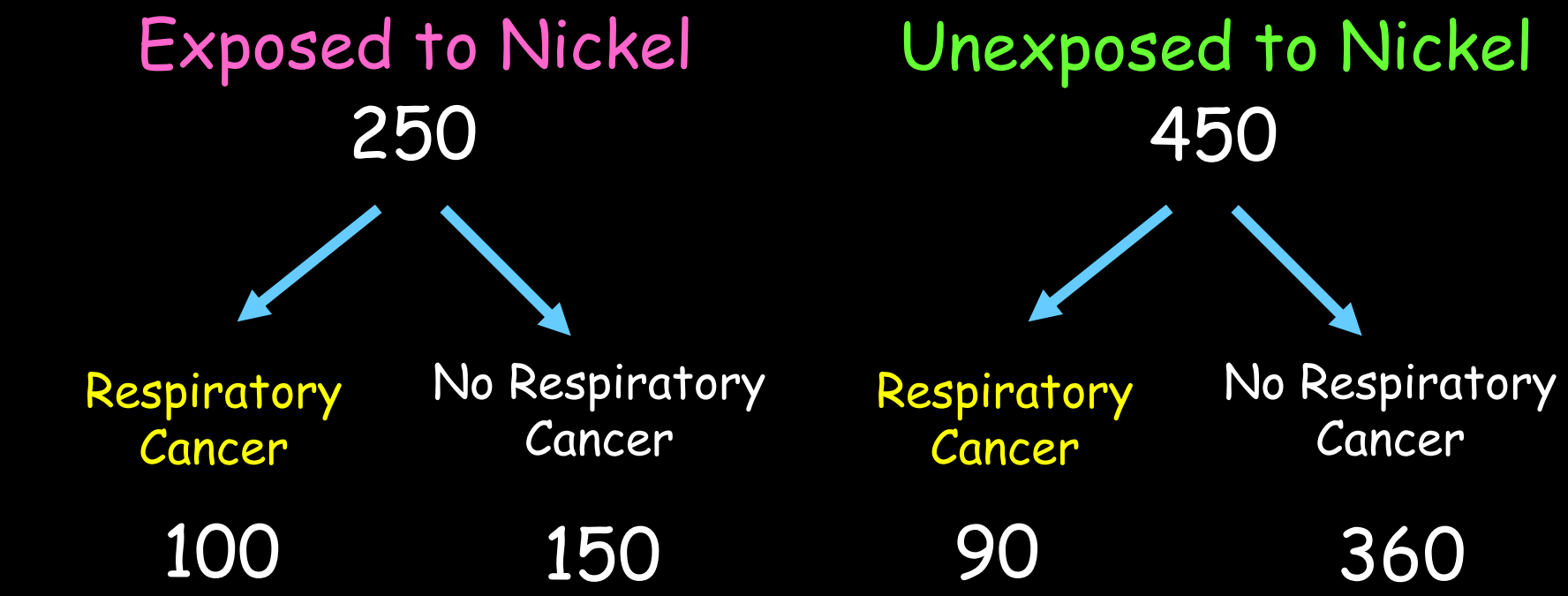
$$\text{Risk}_{47\ \text{yrs}} = \left( \frac{90 \text{ cases of RC}}{450 \text{ subjects free of RC}} \right) = 0.2 = 20\%$$

# Risk in Unexposed

- **Interpretation:**
  - Probability of developing a respiratory cancer in workers unexposed to nickel is 20% over 47 years

# Cohort Design

## SOUTH WALES REFINERY WORKERS



# Risk of respiratory cancer in exposed to nickel

| | Exposed to Nickel |
|---|---|
| Respiratory Cancer | 100 |
| Total | 250 |
| Person-years | 4,100 |

Risk =

**Interpretation:**

# Calculating Risk in Exposed

$$\text{Risk}_{\text{time}} = \left( \frac{\text{New events}}{\text{Population "at risk" at baseline}} \right)$$

$$\text{Risk}_{47\ \text{yrs}} = \left( \frac{100 \text{ cases of RC}}{250 \text{ subjects free of RC}} \right) = 0.4 = 40\%$$

# Risk in Exposed

- **Interpretation:**
  - Probability of developing a respiratory cancer in workers exposed to nickel is 40% over 47 years

# What is an incidence rate of respiratory cancer in this study?

1. Probability of developing RC per 100,000 workers and per year
2. Probability of developing RC over 47 years
3. The excess probability of RC due to exposure
4. The ratio of the probability of disease in exposed over the probability of disease in unexposed
5. Equivalent to the odds of disease (odds of RC)

# Notation

- **R** = Risk
- **IR** = Incidence rate
- **E+** = Exposed to nickel
- **E–** = Non-exposed to dimes
- **R(E+)** = Risk in exposed to nickel
- **IR(E+)** = Incidence rate in exposed to nickel

# Incidence rate (IR) = risk per unit of time

- Risk period = 47 yrs.
- Some subjects followed-up for < 47 yrs.
  - E.g., cases, losses to follow-up

## Solution # 1

- = divide risk by average duration of follow-up (24yrs)

$$\textbf{Risk} = \left( \frac{\text{New RC cases}}{\text{Pop. at risk}} \right)$$

$$\textbf{Incidence Rate} = \left( \frac{\text{New RC cases}}{\text{Pop. at risk} * \text{Duration}} \right)$$

$$\textbf{IR (E–)} = \left( \frac{\text{90 cases RC}}{\text{450 men} * \text{24 yrs}} \right)$$

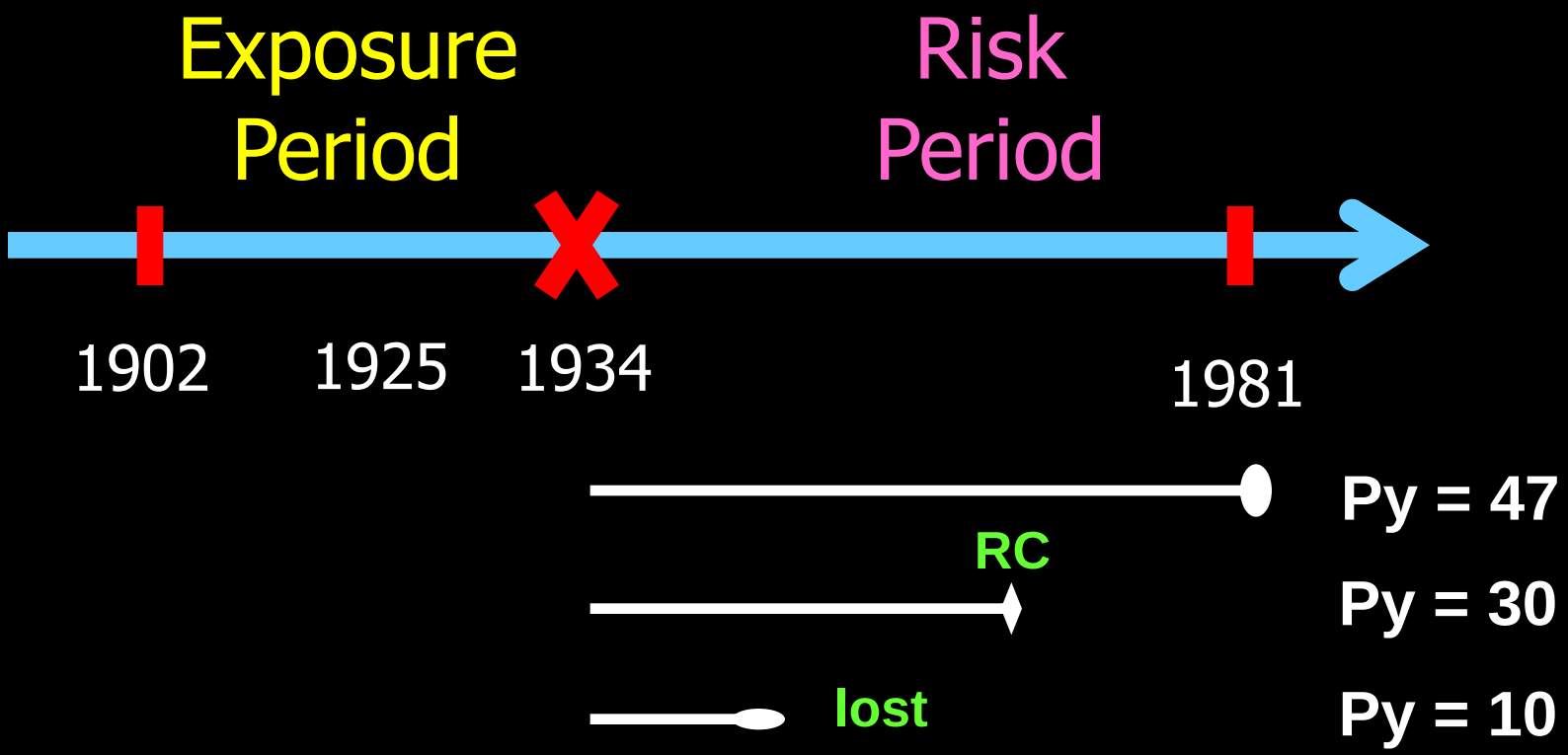
$$= \left( \frac{90}{\text{11,000 person-years}} \right) = \text{0.008/yr}$$

# Incidence rate (IR) = risk per unit of time

## Solution # 2

- Use person-time as denominator
- 1 person followed for 2 years = 2 person-year
- 1 person followed for 1 year = 1 person-year

# Study design



# Example

| | Exposed to Nickel | Unexposed to Nickel |
|---|---|---|
| Respiratory Cancer | 100 | 90 |
| Person-years | 4,100 | 11,000 |
| Incidence Rate | ? | 0.008 |

$$\text{IR (E+)} = \left( \frac{\text{100 cases RC}}{\text{4,100 person-years}} \right)$$

$$= \quad \mathbf{0.024/yr}$$

# What is an attributable risk in this study?

1. The ratio of the risk of RC in exposed to Nickel over the risk in unexposed?
2. The risk of RC that is not due to Nickel exposure
3. The excess rate of RC observed in subjects exposed to nickel compared to unexposed
4. The number of workers that need to be exposed to nickel in order to observe an additional case of RC
5. All of the above

# Absolute Effect:
## Attributable Risk (AR) (2)

AR = IR(E+) - IR(E-)

= IR (E+) - IR (E -)

= 0.024/yr - 0.008/yr = 0.016/yr

= 16 /1,000/y

= Excess IR of RC due to nickel

# Attributable Risk

$$IR(E+) = \left[IR(E-) + AR\right] = \left[0.008 + 0.016\right] = 0.024$$

- Synonymous:
  - Excess Risk
  - Risk Difference
  - Excess Rate

# What is a relative risk in this study?

1. The ratio of the IR of RC in exposed to nickel over the IR in unexposed?
2. The IR of RC that is not due to nickel exposure
3. The excess risk of RC observed among subjects exposed to nickel
4. The number of workers that need to be exposed to nickel in order to observe an additional case of RC
5. None of the above

# Relative Effect: Relative Incidence Rate (RIR)*

$$\mathbf{RIR} = \left( \frac{IR(E+)}{IR(E-)} \right) = \left( \frac{0.024}{0.008} \right) = \mathbf{3.0}$$

* Also referred to as relative risk

# Relative Effect

- Risk in exposed is a multiple of risk in unexposed
- **IR(E+)** = [ IR(E-) * RIR ] = [ 0.008 * 3.0 ]

  = **0.024/yr**

# Relative Effect

**RIR > 1** ..... Nickel exposure increases RC risk

**RIR = 1** ..... No effect of nickel exposure

**RIR < 1** ..... Nickel exposure protects from RC

# Relative or Absolute Effect

| IR(E+) | IR(E–) | RR | AR |
|---|---|---|---|
| 24 /1000/yr | 8 /1000/yr | 3.0 | 16 /1000/yr |
| 60 /1000/yr | 20 /1000/yr | 3.0 | 40 /1000/yr |

# Interpretation

- **Attributable risk** measures clinical and public health importance of the causal relationship
- **Relative risk** assesses strength of the association

# Example: Wrapping up

| | 250 Exposed To Nickel | 450 Unexposed To Nickel |
|---|---|---|
| Respiratory Cancer | 100 | 90 |
| Person-years | 4,100 | 11,000 |
| Incidence Rate | 0.024/yr | 0.008/yr |
| Relative Incidence rate | 3.0 | |
| Attributable Risk | 0.016/yr | |

# Prospective Studies: Advantages

- Exposure to postulated cause is assessed before occurrence of disease
- Possible to estimate all measures of incidence and effect
- Possible to study several outcomes to one cause

# Prospective Studies: Disadvantages

- Requires large investments in time, human and financial resources
- Requires large sample sizes (e.g., 110.000 nurses, 59.600 doctors, 1.2 millions volunteers)
- Not easy to reproduce (Re: consistency of the association)