

Volkert Siersma

Statistiek in de Deense gezondheidszorg

# GROTE DATABASES GEKOPPELD

# Forskningsenheden



# Mijn Ziektekostenverzekeringsbewijs

**REGION H** Region Hovedstaden

ISHØJ KOMMUNE  
Tlf. 43 57 75 75

Læge:  
PER SCHIWE NIELSEN  
VEJLEBROVEJ 84,2. Tlf. 43 73 72 10

060573 XXXX Sikr. gr. 1 Gyldig fra/Valid from:  
22.05.2010


VOLKERT DIRK SIERSMA  
JÆGERBUEN 22,2 TV  
2635 ISHØJ

Certificado de Salud

Tourist health insurance card

SUNDHEDSKORT

REJSESYGESIKRINGSKORT



# CPR nummer

- ⦿ Centraal PersoonsRegistratie nummer
- ⦿ Wordt gebruikt in alle kontakten met:
  - Het rijk
  - De belasting
  - De gemeente
  - De politie
  - De gezondheidszorg
- ⦿ Deze kontakten worden opgeslagen in verscheidene databases

# CPR nummer

- Geïntroduceerd in 1968
- Cijfers 1-6: de geboortedatum (ddmmyy)
- Cijfer 7: de geboorte eeuw:

	Jaar		
Cijfer 7	00-36	37-57	58-99
0-3	1900-1999		
4	2000-2036	1937-1999	
5-8	2000-2057		1858-1899
9	2000-2036	1937-1999	

- Cijfer 10: geslacht (vrouw=even, man=oneven)

# CPR nummer Modulo 11

- Modulo 11 check

Cijfer nr:	1	2	3	4	5	6	7	8	9
Vermenigvuldigen met:	4	3	2	7	6	5	4	3	2

- De som van de producten en cijfer 10 moet deelbaar zijn door 11
- Er zijn ca. 540 CPR nummers voor elke dato
- Vanaf oktober 2007 worden ook niet-Modulo 11 CPR nummers uitgegeven

# De Deense databases

- ◎ [www.ecreph.org](http://www.ecreph.org)
- ◎ Scandinavian Journal of Public Health, Volume 39, Supp. 7, July 2011
- ◎ Denemarken is in een unieke positie om volksgezondheidsonderzoek te doen, omdat het mogelijk is gezondheids- en sociale databases te combineren op individueel niveau via een uniek persoonsidentificatienummer dat gebruikt wordt in alle databases

# Deens civiel registratie systeem

- ◎ Centrale database, sinds 1968
  - Naam
  - Adres
  - Ouders
  - Nationaliteit
  - Dood en emigratie



# Nationale patient database

- ⦿ Ziekenhuiskontakten, sinds 1977
  - Diagnose
  - Operatiecodes
  - Inschrijvingsdatum
  - Uitschrijvingsdatum

# Gezondheidsservice database

- Kontakten met gezondheidsspecialisten buiten ziekenhuiskontakten, sinds 1999
  - Specialist nummer
  - Huisarts
  - Service
    - Basis (huisartsbezoek, etc.)
    - Extra (bloedsample, psychometrie, etc.)
    - Laboratorium test (CRP, etc.)
- Géén diagnoses of test resultaten!

# Kankerregister

- ⦿ Kankerdiagnoses, sinds 1942
  - Diagnose
  - Diagnosedatum
  - Morfologie/stage

# KPLL database

- Kopenhagen huisartsen laboratorium (KPLL) doet alle laboratorium tests voor de huisartsen in (groter) Kopenhagen
- 2000-2010 volledig bloedonderzoek (zonder methodebreuk)
- Volledig bloedonderzoek: het bepalen van het andeel van de verschillende witte bloed lichaampjes in het bloed

# Eosinofilie

- ⦿ Eosinofilie: een verhoogd aantal eosinofiele granulocyten
- ⦿ Wit bloedlichaampje tegen parasieten (meer-cellige) infecties
- ⦿ ...tegen astma en allergie
  
- ⦿ ...maar kan schadelijk zijn, wellicht een verhoogde kans op bloedkanker!

# KPLL database

- ◎ 1,579,524 proeven van 621,547 personen
  - Verwijder personen met ongeldige CPR
  - Verwijder personen die niet in het civiel registratiesysteem voorkomen
- ◎ 1,573,012 proeven van 616,540 personen
  - Enkel bloedproeven aangevraagd door huisartsen
- ◎ 1,299,945 proeven van 545,505 personen

# KPLL database

- ⦿ Één toevallig uitgekozen proef per persoon
  - Tussen 18 en 80 jaar oud
  - In de periode 2001-2007
- ⦿ 359,950 proeven en personen
- ⦿ 356,027 proeven met een bruikbare meting van eosinofilie

# Bloedkanker

- ⦿ Haematologische kanker
- ⦿ Bepaalde diagnose codes (ICD-10) uit het kanker register
- ⦿ Wij bekijken de voorkomst van bloedkanker binnen drie jaar vanaf de datum van de toevallig uitgekozen bloedproef
- ⦿ Personen die al een bloedkankerdiagnose hebben vóór deze datum worden uit de analyse gehaald.



# Eosinofilie en bloedkanker

	Kanker < 3 jaar		
Eosinofilie	Nee	Ja	Totaal
Nee	341174	462	341636
Ja	14346	45	14391
Totaal	355520	507	356027

# Eosinofilie en bloedkanker

	Kanker < 3 jaar			
Eosinofilie	Nee	Ja	Totaal	Risico (p)
Nee	341174	462	341636	0.14%
Ja	14346	45	14391	0.31%
Totaal	355520	507	356027	

$$\begin{aligned} p_{\text{Nee}} &= 462:341636 &= 0.0014 &= 0.14\% \\ p_{\text{Ja}} &= 45:14391 &= 0.0031 &= 0.31\% \end{aligned}$$

# Eosinofilie en bloedkanker

	Kanker < 3 jaar			
Eosinofilie	Nee	Ja	Totaal	Risico (p)
Nee	341174	462	341636	0.14%
Ja	14346	45	14391	0.31%
Totaal	355520	507	356027	

**Relatief risico (RR) =  $0.0031/0.0014 = 2.21$**

# Eosinofilie en bloedkanker

	Kanker < 3 jaar			
Eosinofilie	Nee	Ja	Totaal	Risico (p)
Nee	341174	462	341636	0.14%
Ja	14346	45	14391	0.31%
Totaal	355520	507	356027	

Relatief risico (RR) =  $0.0031/0.0014 = 2.21 \approx 1?$

# Relatief risico (RR)

- ⦿ Een relatief risico is een getal  $\geq 0$
- ⦿ In de epidemiologie wordt dit ook wel een relatieve risico reductie (RRR) genoemd
- ⦿ Als we mensen zouden behandelen voor eosinofilie zouden we het aantal kankers kunnen halveren ( $45 \rightarrow 23$ )
- ⦿ Een relatief risico is ziet er vaak dramatisch uit in verhouding tot de risico's

# Eosinofilie en bloedkanker

	Kanker < 3 jaar			
Eosinofilie	Nee	Ja	Totaal	Risico (p)
Nee	341174	462	341636	0.14%
Ja	14346	45	14391	0.31%
Totaal	355520	507	356027	

$$\text{Risicoverschil (RD)} = 0.0014 - 0.0031 = -0.0017$$

# Eosinofilie en bloedkanker

	Kanker < 3 jaar			
Eosinofilie	Nee	Ja	Totaal	Risico (p)
Nee	341174	462	341636	0.14%
Ja	14346	45	14391	0.31%
Totaal	355520	507	356027	

$$\text{Risicoverschil (RD)} = 0.0014 - 0.0031 = -0.0017 \approx 0?$$

# Risicoverschil (RD)

- ⦿ Een risicoverschil ligt tussen -1 en 1
- ⦿ In de epidemiologie wordt dit ook wel een absolute risico reductie (ARR) genoemd
- ⦿ Als we mensen zouden behandelen voor eosinofilie dan  $ARR=0.0017$
- ⦿ We zouden dan  $1:0.0017 = 588$  mensen moeten behandelen om één kanker te voorkomen (NNT)

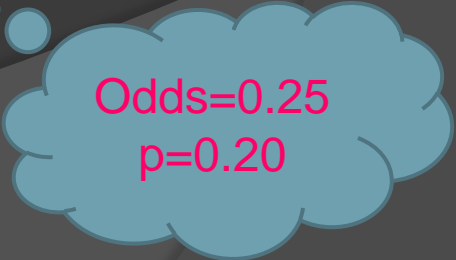


# Odds

- ⦿ Een andere manier om risico's (kansen) uit te drukken
- ⦿ De verhouding tussen de kans op een gebeurtenis en de kans dat die gebeurtenis *niet* optreed
- ⦿ Odds =  $p:(1-p)$
- ⦿ In kansspelen worden odds gebruikt om de win-ratio aan te geven; **1:4** wil zeggen €1 wordt €4 als je wint

# Odds

- ⦿ Een andere manier om risico's (kansen) uit te drukken
- ⦿ De verhouding tussen de kans op een gebeurtenis en de kans dat die gebeurtenis *niet* optreed
- ⦿ Odds =  $p:(1-p)$
- ⦿ In kansspelen worden odds gebruikt om de win-ratio aan te geven; **1:4** wil zeggen €1 wordt €4 als je wint



Odds=0.25  
p=0.20

# Eosinofilie en bloedkanker

	Kanker < 3 jaar			
Eosinofilie	Nee	Ja	Totaal	Odds
Nee	341174	462	341636	0.0014
Ja	14346	45	14391	0.0031
Totaal	355520	507	356027	

$$\text{Odds}_{\text{Nee}} = 462:341174 = 0.0014$$

$$\text{Odds}_{\text{Ja}} = 45:14346 = 0.0031$$

# Eosinofilie en bloedkanker

	Kanker < 3 jaar			
Eosinofilie	Nee	Ja	Totaal	Odds
Nee	341174	462	341636	0.0014
Ja	14346	45	14391	0.0031
Totaal	355520	507	356027	

Odds ratio (OR) =  $0.0031/0.0014 = 2.32$

# Eosinofilie en bloedkanker

	Kanker < 3 jaar			
Eosinofilie	Nee	Ja	Totaal	Odds
Nee	341174	462	341636	0.0014
Ja	14346	45	14391	0.0031
Totaal	355520	507	356027	

$$\text{Odds ratio (OR)} = 0.0031/0.0014 = 2.32 \approx 1?$$

# Odds ratio (OR)

- ⦿ Een odds ratio is een getal  $\geq 0$
- ⦿ Heeft niet een direkte epidemiologische interpretatie
- ⦿ Als  $p$  klein is, dan **odds =  $p:(1-p) \approx p$**
- ⦿ ...en dus **OR  $\approx$  RR**
- ⦿ Dus als we kunnen aannemen dat de ziekte zeldzaam is, dan kunnen we de odds ratio interpreteren als een relatief risico (rare disease assumption)

# Kanker is zeldzaam

- ⦿ In onze steekproef is het aantal kankers klein; de getallen in de “Ja” kolom zijn veel kleiner dan in de “Nee” kolom
- ⦿ Inderdaad zien we dat...
- ⦿  $RR = 2.21 \approx 2.32 = OR$

# Statistische tests

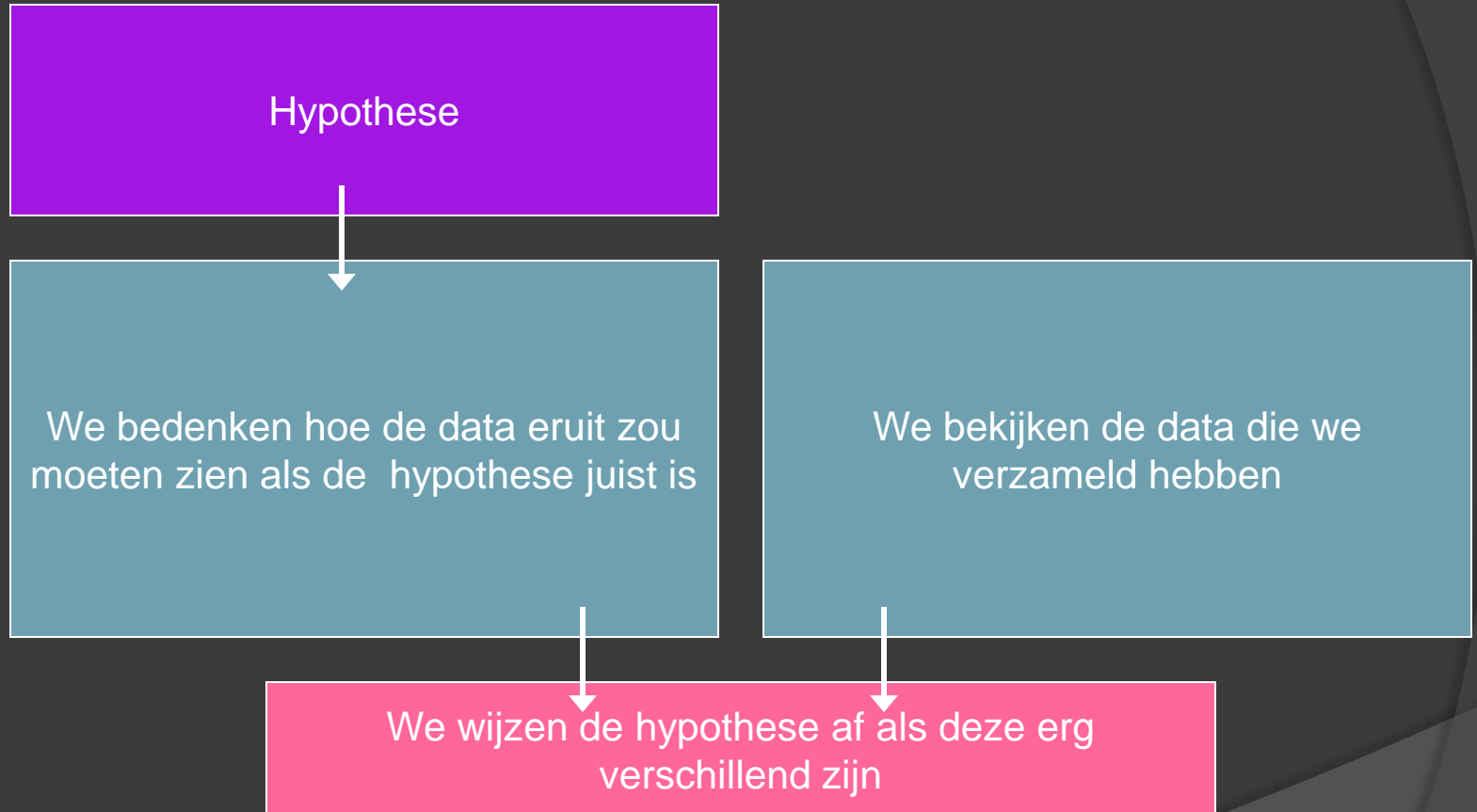
- ⦿ Als er geen verschil in risico is dan zouden we  $OR=1$  moeten vinden
- ⦿ We vinden  $OR=2.32$  in onze steekproef
- ⦿ Natuurlijk vinden we nooit precies  $OR=1$
- ⦿ Is  $OR=2.32$  wezenlijk verschillend van  $OR=1$ ?



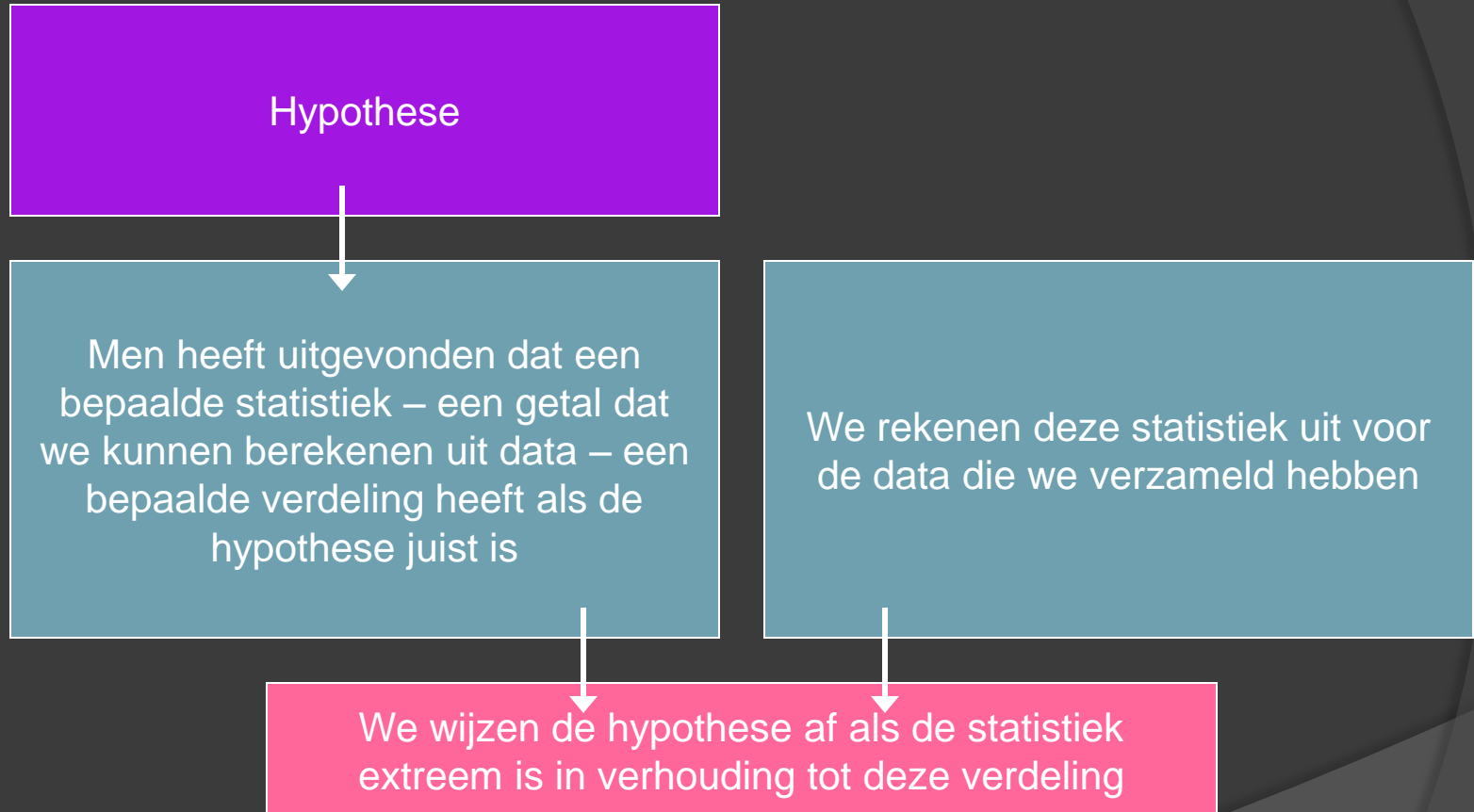
# Statistische tests

- ⦿ Als er geen verschil in risico is dan zouden we  $OR=1$  moeten vinden
- ⦿ We vinden  $OR=2.32$  in onze steekproef
- ⦿ Natuurlijk vinden we nooit precies  $OR=1$
- ⦿ Is  $OR=2.32$  wezenlijk verschillend van  $OR=1$ ?
- ⦿ ...of  $RR=2.21$  verschillend van  $RR=1$ ?
- ⦿ ...of  $RD=-0.0017$  verschillend van  $RD=0$ ?

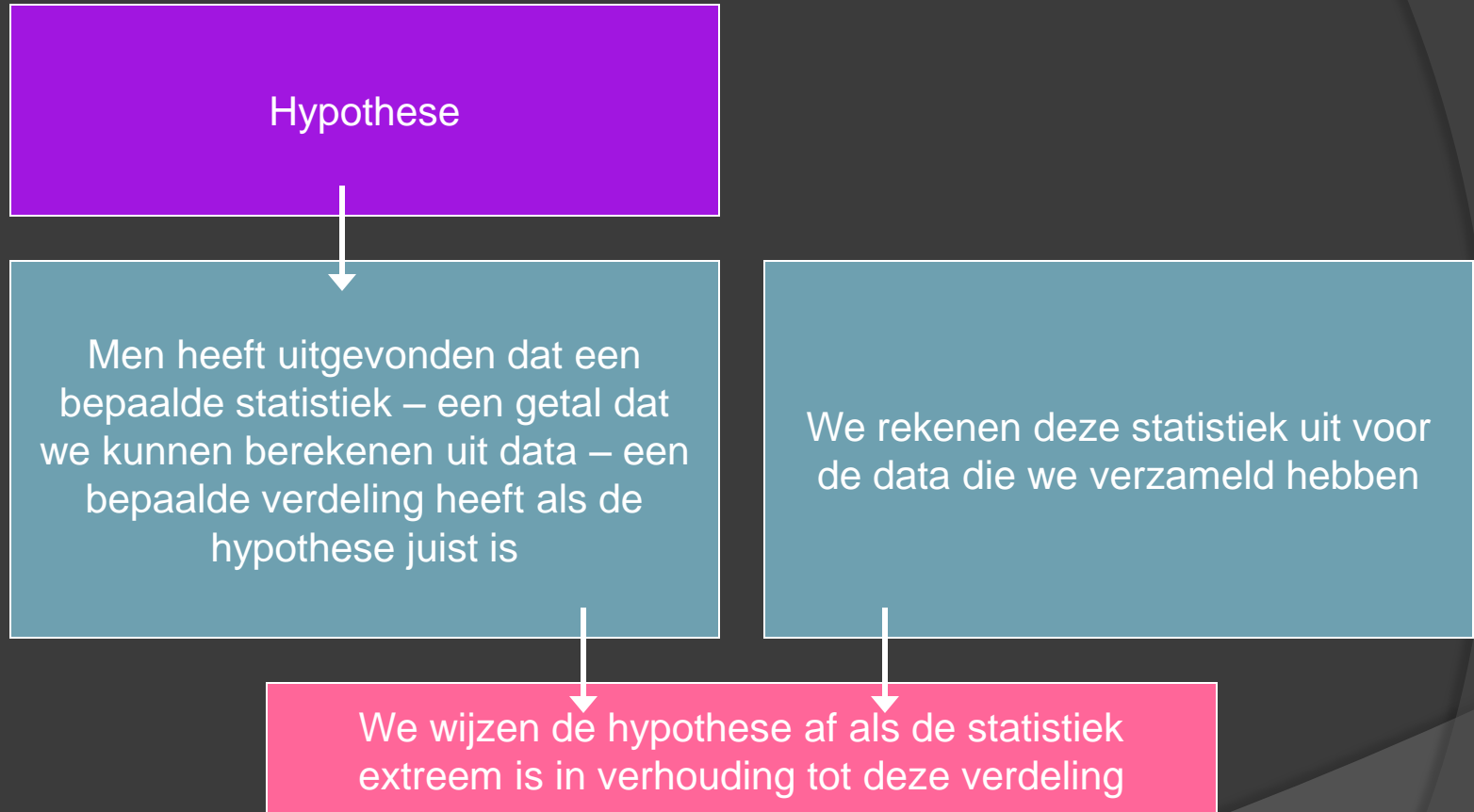
# Statistische tests



# Statistische tests



# Statistische tests



# Statistische tests

- ⦿ We willen de nul-hypothese  $H_0$  afwijzen
- ⦿  $H_0: OR=1 \leftrightarrow \log(OR)=0$
- ⦿ Als we  $H_0$  kunnen afwijzen, dan accepteren we de alternatieve hypothese  $H_1$
- ⦿  $H_1: OR \neq 1 \leftrightarrow \log(OR) \neq 0$

# T-test

- Student's T statistiek

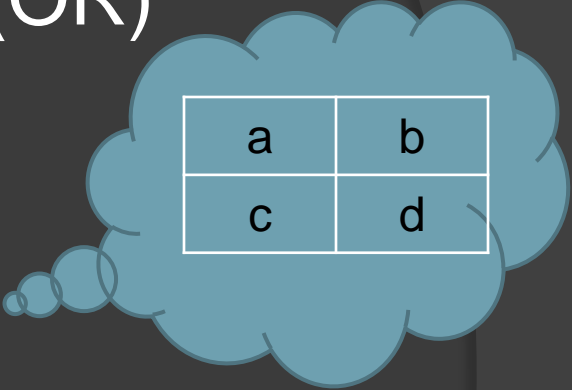
$$T = \frac{\log(\text{OR})}{\text{SE}(\log(\text{OR}))}$$

- Als T groot is, dan wijzen we  $H_0$  af
- Als  $H_0$  juist is en we veel (...) data hebben, dan heeft T een t-verdeling
- Een t-verdeling praktisch hetzelfde als een standard Normale verdeling

# Standaard fout (SE)

- De standaard fout (SE) van  $\log(\text{OR})$  wordt uitgerekend als

$$\text{SE}(\log(\text{OR})) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$



a	b
c	d

- Als we meer data krijgen, dan...
- a, b, c en d  $\uparrow$
- $\text{SE}(\log(\text{OR}))$   $\downarrow$
- T  $\uparrow$

# Statistische tests

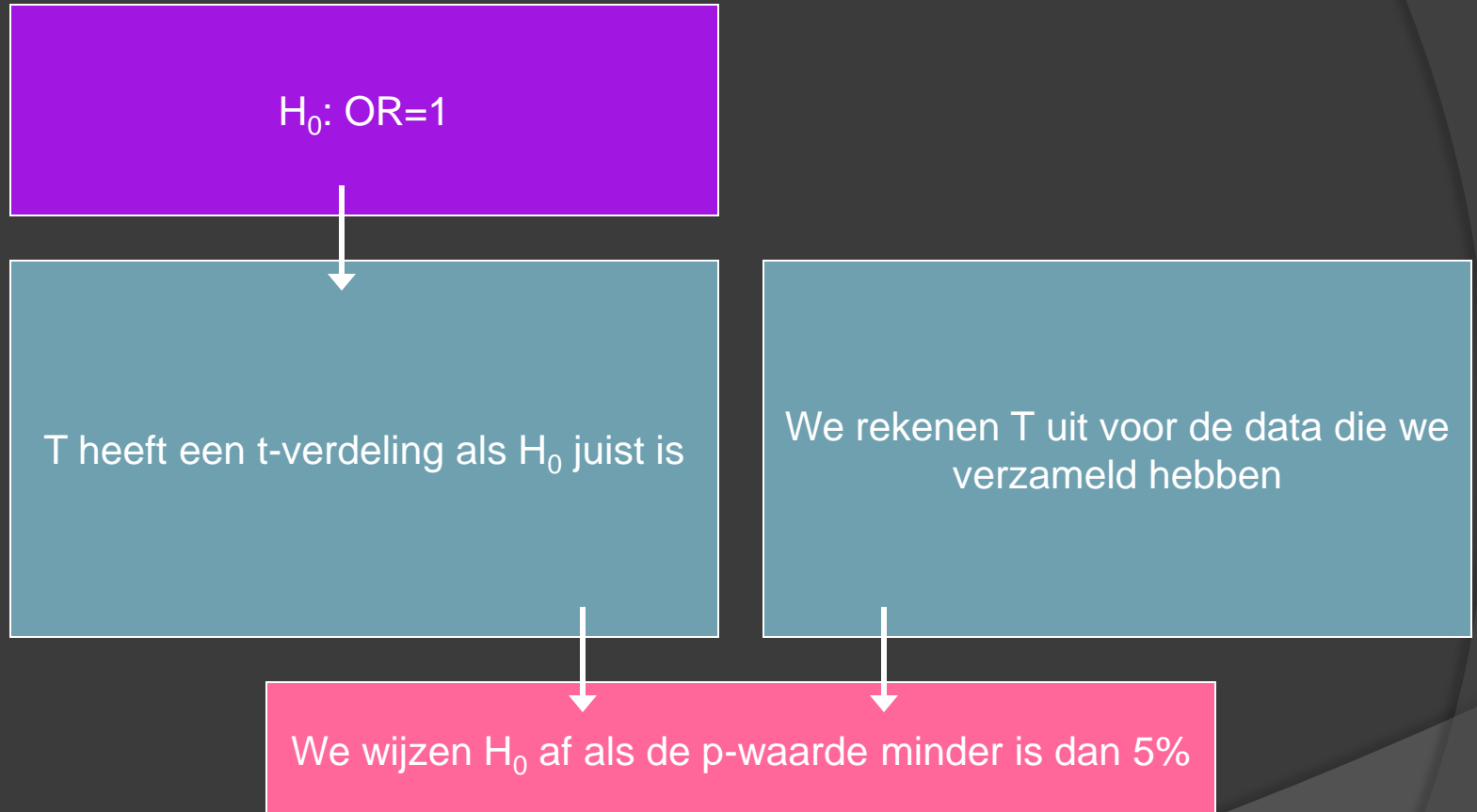




# Wat is extreem?

- ⦿ We wijzen  $H_0$  af als T extreem – positief of negatief – is in verhouding tot een t-verdeling
- ⦿ We zeggen dat T extreem is als de kans dat we T – of iets méér extreems – observeren als  $H_0$  eigenlijk juist is minder dan 5% is
- ⦿ Deze kans wordt ook wel p-waarde genoemd

# Statistische tests



# Eosinofilie en bloedkanker

$$OR=2.31 \rightarrow \log(OR)=0.84$$

$$SE(\log(OR)) = \sqrt{\frac{1}{341174} + \frac{1}{462} + \frac{1}{14346} + \frac{1}{45}} = 0.16$$

$$T = 0.84:0.16 = 5.38$$

$$\begin{aligned} \Pr(T < -5.38 \text{ of } T > 5.38) &= \\ &= 2 \times \Pr(T < -5.38) = \\ &= 2 \times 0.0000000037 < 0.05 \end{aligned}$$

# Grote databases

- ⦿ De kans om een verband te vinden dat er niet is, is 5%
- ⦿ Maar zelfs erg kleine – maar reële – verschillen in risico's worden significant; hebben een p-waarde  $<0.05$
- ⦿ Dankbaar voor de epidemioloog die zijn resultaten gepubliceerd wil zien (...)
- ⦿ Maar het is de vraag of deze verschillen klinisch relevant zijn...

# Case-control studies

- In een klassieke case-control studie worden mensen gesampled naar de ziekte, bijvoorbeeld:
- ...alle mensen die kanker krijgen (cases)
- ...voor elk van dezen een enkele persoon die geen kanker krijgt (controls)
- In het algemeen is het risico op kanker anders (vaak hoger) dan in een representatief sample

# Eosinofilie case-control studie

	Kanker < 3 jaar		
Eosinofilie	Nee	Ja	Totaal
Nee	487	462	949
Ja	20	45	65
Totaal	507	507	1014

## Gefingeerd voorbeeld

de rij-verhouding in de “Nee” kolom is dezelfde als in de volledige steekproef

# Eosinofilie case-control studie

	Kanker < 3 jaar			
Eosinofilie	Nee	Ja	Totaal	Risico (p)
Nee	487	462	949	48.02%
Ja	20	45	65	69.23%
Totaal	507	507	1014	

$$p_{\text{Nee}} = 462:962 = 0.4802 = 48.02\%$$

$$p_{\text{Ja}} = 45:65 = 0.6923 = 69.23\%$$

# Eosinofilie case-control studie

- ⦿ RR = 1.44 (was 2.21)
- ⦿ RD = -0.2121 (was -0.0017)



# Eosinofilie case-control studie

	Kanker < 3 jaar			
Eosinofilie	Nee	Ja	Totaal	Odds
Nee	487	462	949	0.9487
Ja	20	45	65	2.2500
Totaal	507	507	1014	

$$\text{Odds}_{\text{Nee}} = 462:487 = 0.9487$$

$$\text{Odds}_{\text{Ja}} = 45:20 = 2.2500$$

# Eosinofilie case-control studie

- ◎ RR = 1.44 (was 2.21)
- ◎ RD = -0.2121 (was -0.0017)
- ◎ OR = 2.37 (was 2.32)

# Eosinofilie case-control studie

- ◎ RR = 1.44 (was 2.21)
- ◎ RD = -0.2121 (was -0.0017)
- ◎ OR = 2.37 (was 2.32) !!!
  
- ◎ ...dat de OR van de case-control studie niet **exact hetzelfde** is als de OR van de volledige steekproef komt omdat in het gefingeerde voorbeeld afgerond is naar gehele getallen

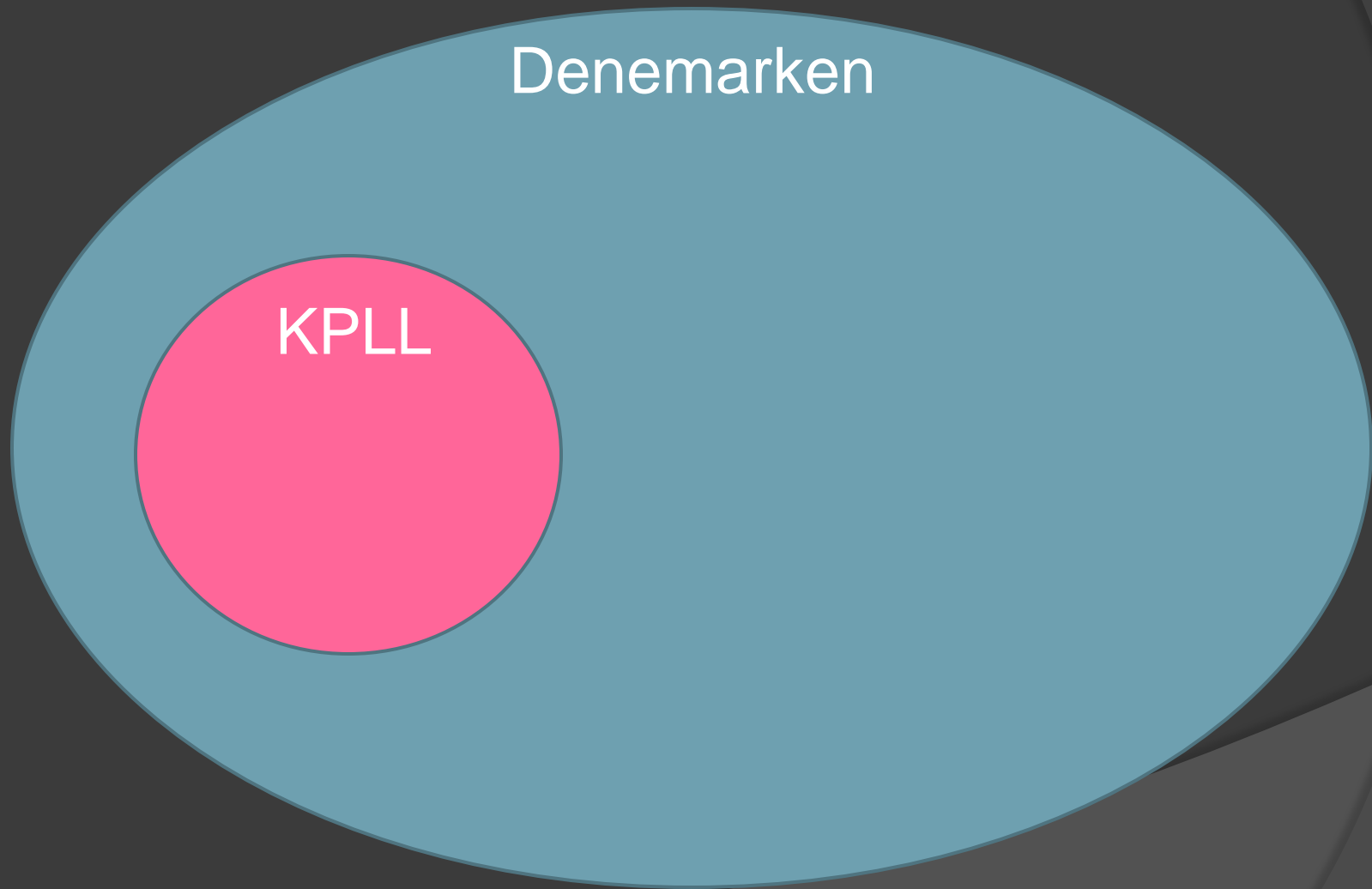
# Odds ratio

- ⦿ De odds ratio is onafhankelijk van de marginale (rij- en kolom-) verdelingen van de tabel
- ⦿ Daarbij is de odds ratio dus onafhankelijk van de manier waarop de steekproef werd geselecteerd

# De KPLL database

- ⦿ De personen in de KPLL database zijn **niet** representatief voor de Deense bevolking
- ⦿ Er was een reden dat de huisarts een bloedproef liet doen
- ⦿ De KPLL database kan dus worden gezien als een groep met een risico op kanker dat anders (hoger?) is dan in de Deense bevolking in het algemeen

# De vertaling van risico's



# De vertaling van risico's

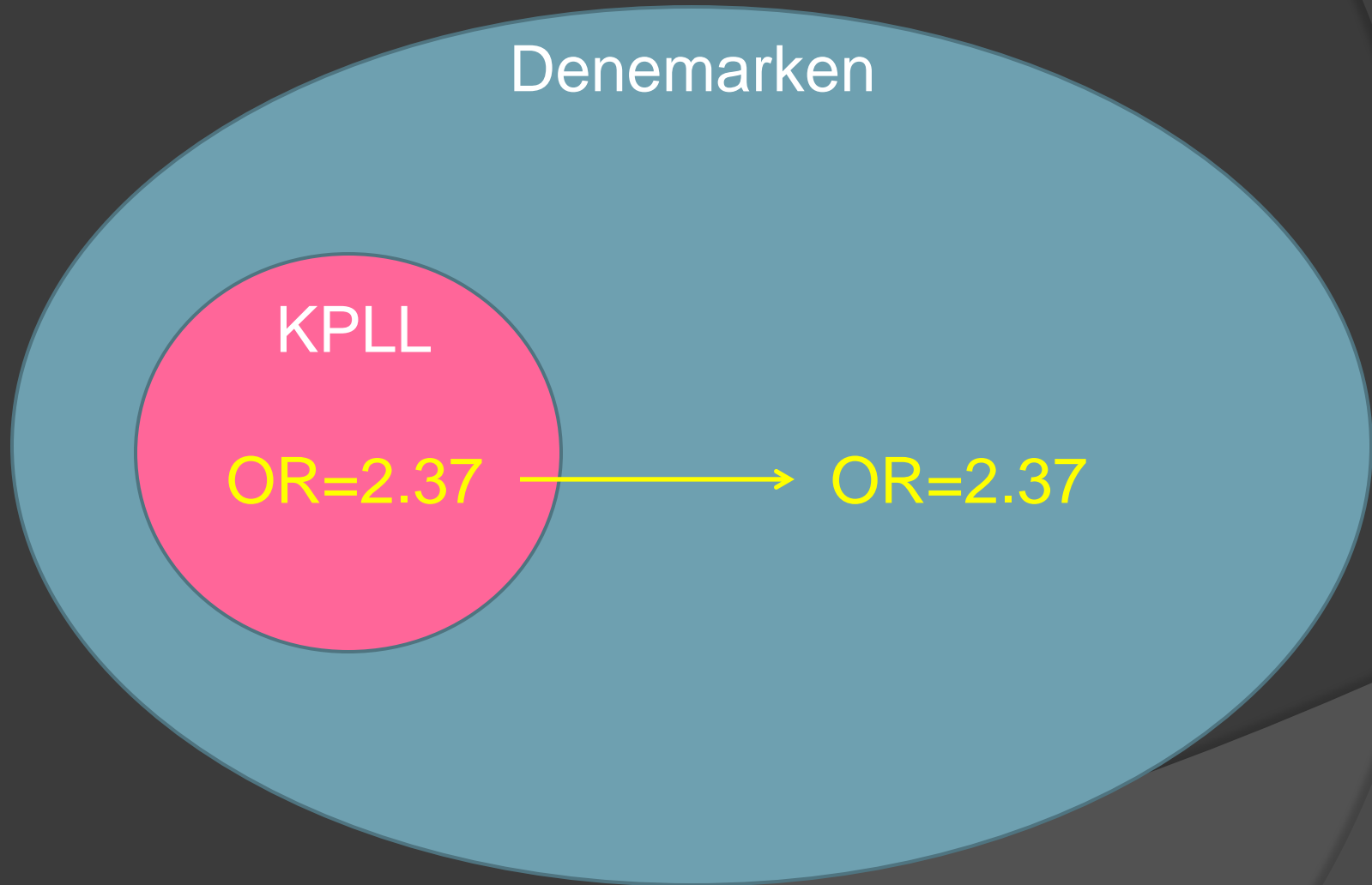
Denemarken



KPLL

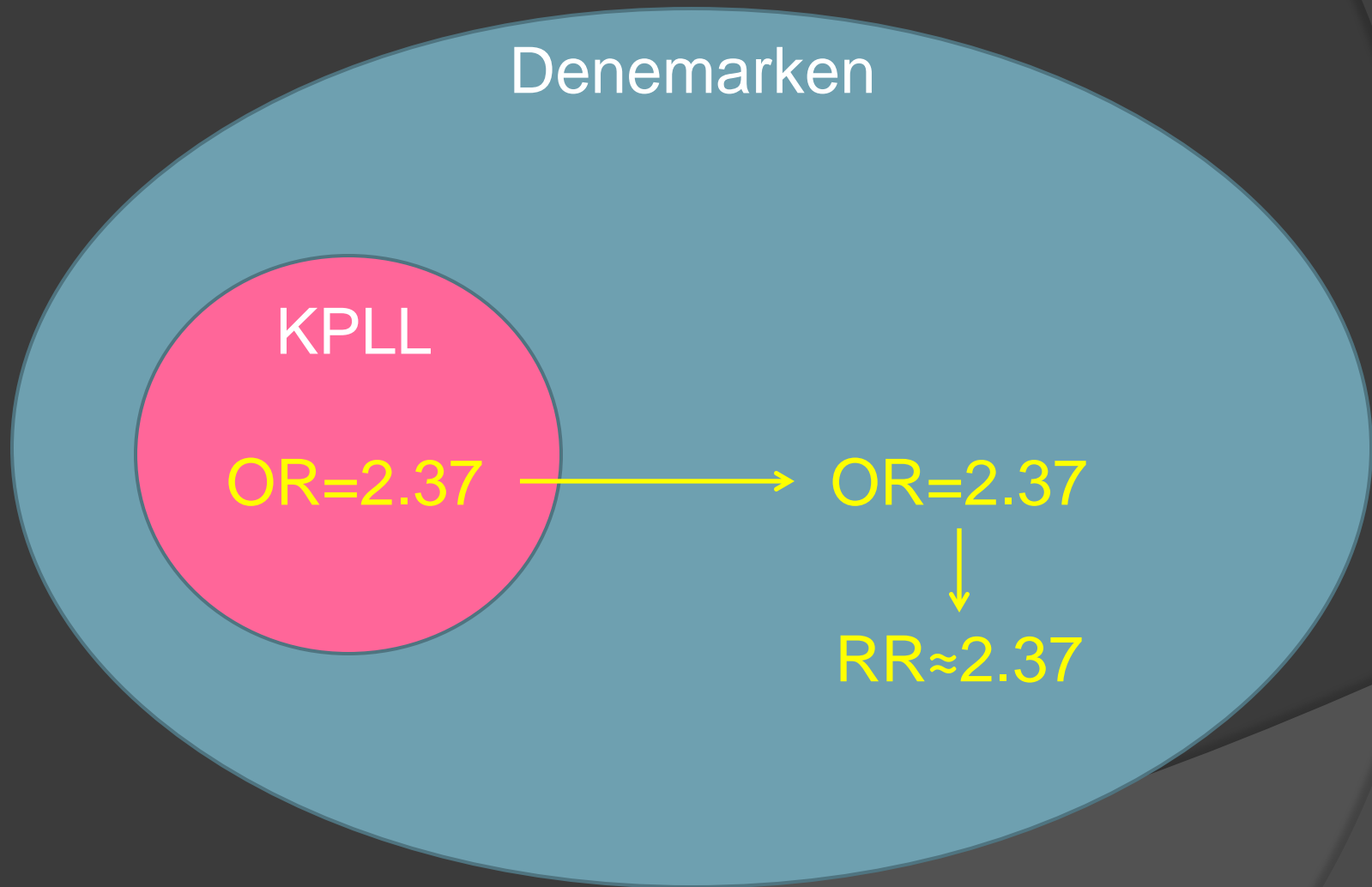
OR=2.37

# De vertaling van risico's





# De vertaling van risico's



# De vertaling van risico's

Denemarken

Eosinofilie méér dan verdubbelt  
de kans op bloedkanker