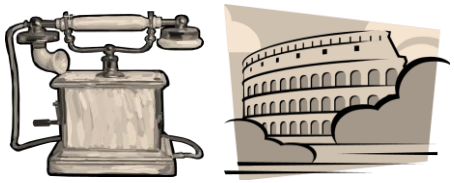
Tele-ARENA

(**Tele**fonisch **A**angenaam **R**efereren **E**n **N**etwerken **A**MPHI)



3. Data-analyse

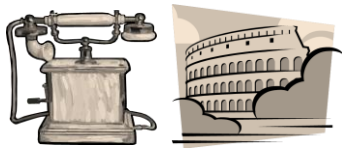
Van dataverzameling naar resultaat

Alma Tostmann

Epidemioloog @ Academische Werkplaats AMPHI

Afdeling Eerstelijnsgeneeskunde, UMC St Radboud, Nijmegen

1 oktober 2013



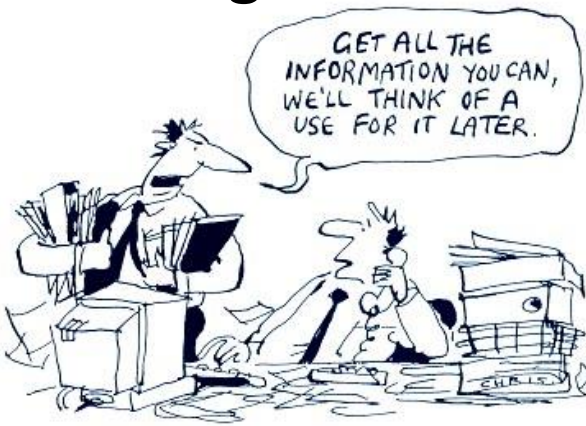
Welkom bij de 3e Tele-ARENA

- Tele-ARENA = telefonische nascholing opzet wetenschappelijk onderzoek
- Onderwerpen + praktijkvoorbeeld:
  1. **Het onderzoeksvoorstel**: van onderzoeksvraag naar doelstellingen
  2. **Dataverzameling**: van doelstellingen naar dataverzamelingsplan
  3. **Data analyse**: beschrijvende & statistische analyse – een opfrisser
  4. **Onderzoeksrapportage**: data presentatie, kernboodschap, aanbevelingen

Tijdens de presentatie:

- Start: deelnemers zetten telefoon op **luisterstand: intoetsen "16"**
- Vragen tussendoor mag (eerst telefoon uit luisterstand zetten: 16)

Wrap-up 2e Tele-ARENA - dataverzameling



Source: http://www.priv.gc.ca/information/ar/images/cartoon3.jpg

- Maak een analyseplan vóór aanvang dataverzameling
- Analyse is gestuurd door vraagstelling
- Dataverzameling gestuurd door vraagstelling &analyseplan

Want je wilt:

- Voldoende data verzamelen om je vraag te beantwoorden
- Maar, niet meer data dan nodig.

Aan de hand van de onderzoeksvraag bepaal je:

  - **Primaire uitkomstmaat** (wat wil je meten)
  - **Determinanten** (bv groepen waartussen je wilt vergelijken)
  - Mogelijk **verstorende variabelen** (confounding/ effect modificatie)
- Meetbaar maken, welke databronnen, hoe samenvoegen en verwerken
- Altijd op het netvlies houden:
  - Wat is de onderzoeksvraag? (voorkomen dat je data vergeet of teveel verzamelt)
  - Wat is het analyseplan (zorg dat de data in de juiste vorm verzameld wordt)

Programma vandaag

“De data is verzameld, en dan?”

1. Verschillende typen variabelen: wat doe je ermee
2. De analyse: eerst beschrijven, dan het effect en de statistische test
3. Voorbeeld uit de praktijk

Data is verzameld, en dan?

- Zorg dat alle data bij elkaar staat, niet perse in 1 database.
- Terug naar onderzoeksvraag en analyseplan, daar staat:
  - Welk **getal** gaat antwoord geven op de vraag en hoe kom je aan dat getal?
  - Dus: wat is de primaire uitkomstmaat & determinanten
- Welk type variabelen heb ik mee te maken? (*volgende slide*…)
- Wat is de relatie tussen de groepen die ik vergelijk?
  - Onafhankelijke groepen: twee groepen, bv interventie & controle, ziek & niet ziek
  - Afhankelijk: voor-na meting in dezelfde groep, twee groepen zijn aan elkaar gematcht
    Gaan we niet op in vandaag: uitgaan van onafhankelijke groepen, komt het meest voor.
- Waarmee analyseren:
  - Achterkant bierviltje? Bij simpele analyse moet dit in principe kunnen ☺
  - Excel: simpele overzichten maken kan wel; programma niet bedoeld voor data-analyse
  - Statistisch programma (SPSS, STATA) en/of hulp epidemioloog & statisticus
  - Online tools: bv EpiSheet (zie link op laatste pagina)

Verschillende typen variabelen

- Het **type variabele** bepaald het soort beschrijvende en statistische analyse je ermee kunt (moet) uitvoeren.

Type variabelen	Kenmerk	Voorbeeld
Interval (continu)	Op een continue schaal te meten	Leeftijd, gewicht, bloeddruk
Categorisch nominaal:	Categorieën zonder ordening	
Dichotoom	2 groepen, ongeacht ordening	Geslacht m/v, ziek ja/nee.
>2 groepen	Meerdere groepen, geen ordening	Nationaliteit (A, B, C)
Categorisch ordinaal	Categorieën met duidelijke ordening, maar niet persé zelfde afstand	Ernst symptomen 0, I, II, III

- Beschrijving: meestal gemiddelde (met SD) of een aantal (met percentage)
- Effectmaat: meestal een verschil of een relatief risico
- Statistiek: Meestal T-test of chi-kwadraat

Twee voorbeelden

Voorbeeld 1: tuberculoseonderzoek uit Tanzania

INT J TUBERC LUNG DIS 13(12):1524–1529
©2009 The Union

Community vs. facility-based directly observed treatment for tuberculosis in Tanzania's Kilimanjaro Region

J. van den Boogaard,* R. Lyimo,† C. F. Irongo,‡ M. J. Boeree,*§ H. Schaalma,¶† R. E. Aarnoutse,* G. S. Kibiki†

*Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; †Kilimanjaro Christian Medical Centre, Moshi, ‡National TB and Leprosy Programme, Kilimanjaro Region, Tanzania; §University Lung Centre Dekkerswald, Groesbeek, ¶Maastricht University, Maastricht, The Netherlands

Behandeluitkomst 'succesvol' (is *behandeling afgemaakt* of *genezen*), 'niet succesvol' (is o.a. *overleden*, *gestopt*, *lost-to-follow-up*)

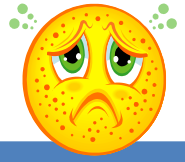
Directly Observed Treatment (DOT): strategie waarbij o.a. observer is bij het innemen van de medicatie. Kan in het gezondheidscentrum, of thuis.

Voorbeeld 2:

Mening t.a.v. invoer varicellavaccinatie in RVP onder verpleegkundigen en artsen op consultatiebureau & jeugdgezondheidszorg.

1 – TBC in Tanzania

Probleem	**Gunstige behandeluitkomst bij ongeveer 77% van TB patiënten in noord Tanzania. Wat zijn voorspellers voor goede uitkomst?**
1. Onderzoeksvraag	Verhoogt het monitoren van tuberculosebehandeling in het ziekenhuis het succesvol afronden van de TBC behandeling vergeleken met thuismonitoring?
Setting / dataverzameling	*Data van Nationaal TB programma Tanzania: alle records uit 2009 uit de Kilimanjaro Region*
2. Primaire uitkomstmaat	Percentage TB patiënten met 'succesvolle' behandeluitkomst *Dus nodig:* *Informatie van alle afzonderlijke TB patiënten (totaal aantal)* *Behandeluitkomst van al die patienten (% succesvol / totaal*)
3. Indicatoren/ determinanten	Plaats monitoring (ziekenhuis of thuis); HIV status; Leeftijd/geslacht *Dus nodig:* *Uitslag HIV test, geslacht, leeftijd (waarschijnlijk alleen categorie < of > 15 jaar)*
4. Type variabelen	Primaire uitkomstmaat: dichotoom (succesvolle uitkomst ja/nee) Determinanten = allemaal dichotoom; *Data uit TB registratie, alleen <15 jaar of >= 15 jaar.*



2 – varicella vaccinatie & RVP

Probleem	Peilen mening gezondheidswerkers jeugd t.a.v. varicellavaccinatie in het RVP?
1. Onderzoeksvraag	Wat is de mening t.a.v. invoer van varricellavaccinatie in het RVP en welke factoren zijn geassocieerd met een positieve mening?
Setting / dataverzameling	*Online vragenlijst verstuurd naar consultatiebureau & JGZ medewerkers; in 2012.*
2. Primaire uitkomstmaat	Percentage medewerkers dat 'voor' invoer v.vaccinatie in RVP is. *Dus nodig:* *Mening van elke geïnterviewde; in dit geval was het een antwoord op een vraag over die mening.*
3. Indicatoren/ determinanten	Kennis over waterpokken, mening over veiligheid vaccin, opvatting ernst waterpokken, werkplek (JGZ of IZB), arts of vp. *Dus nodig:* *Kennis: totale score op 7 vragen (vervolgens opgesplitst in goed/matig), mening, werkplek en beroep*
4. Type variabelen	Primaire uitkomstmaat: dichotoom (voor of tegen) Determinanten = meesten dichotoom, of categorisch (likert scale)

Deel 2: Analyse
Vier situaties:

Tussen groep A en groep B wil je vergelijken:

1. Continue variabele
2. Dichotome variabele (=2 categorieën)
3. Categorische variabele (>2 categorieën)

Binnen een groep wil je vergelijken:

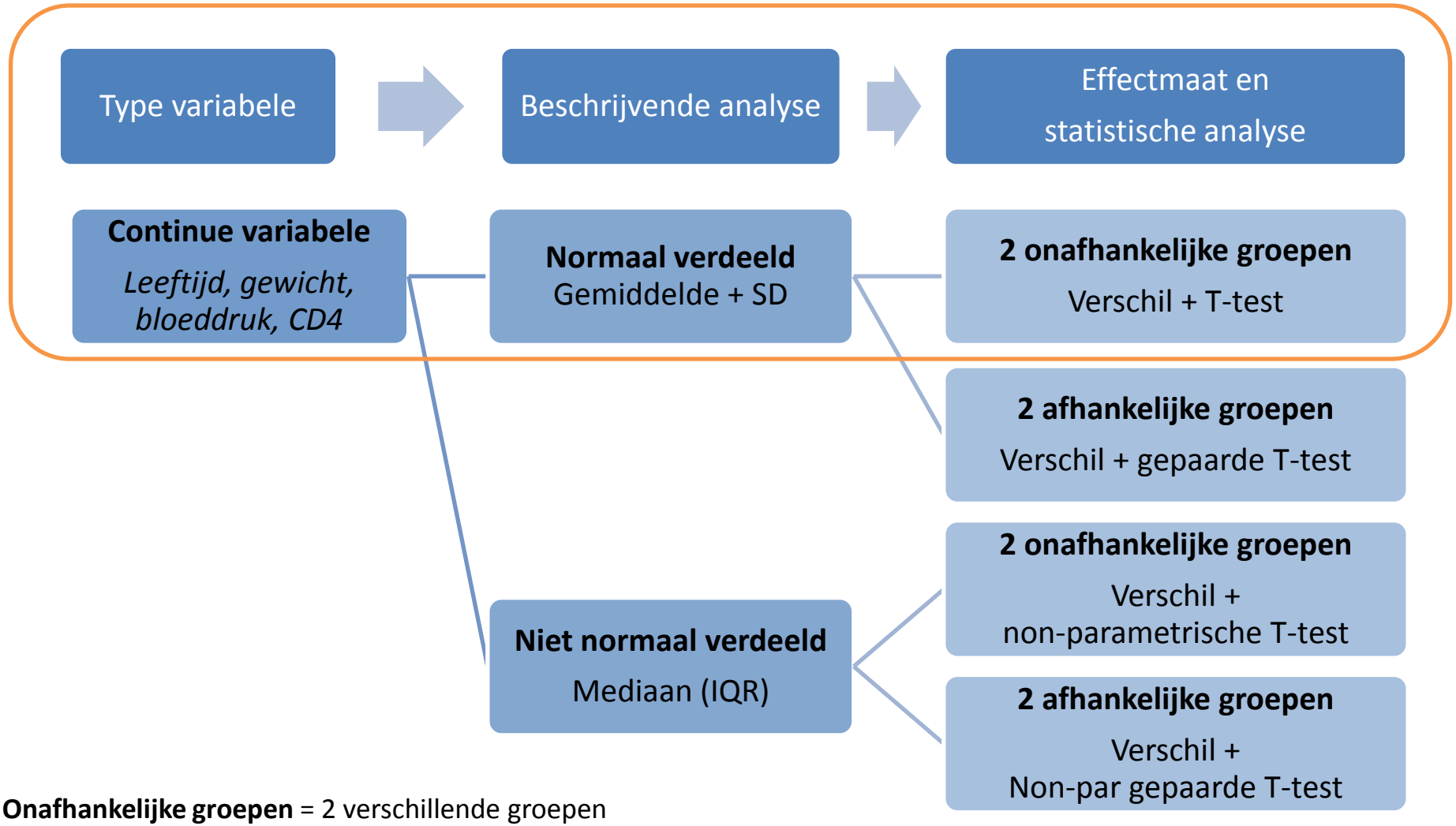
4. Twee continue variabelen)

Maak onderscheid tussen: (NB alle 3 even belangrijk)

1. Beschrijving
2. De effectmaat
3. De statistische test die een p-waarde geeft

Beschrijvende en statistische analyses

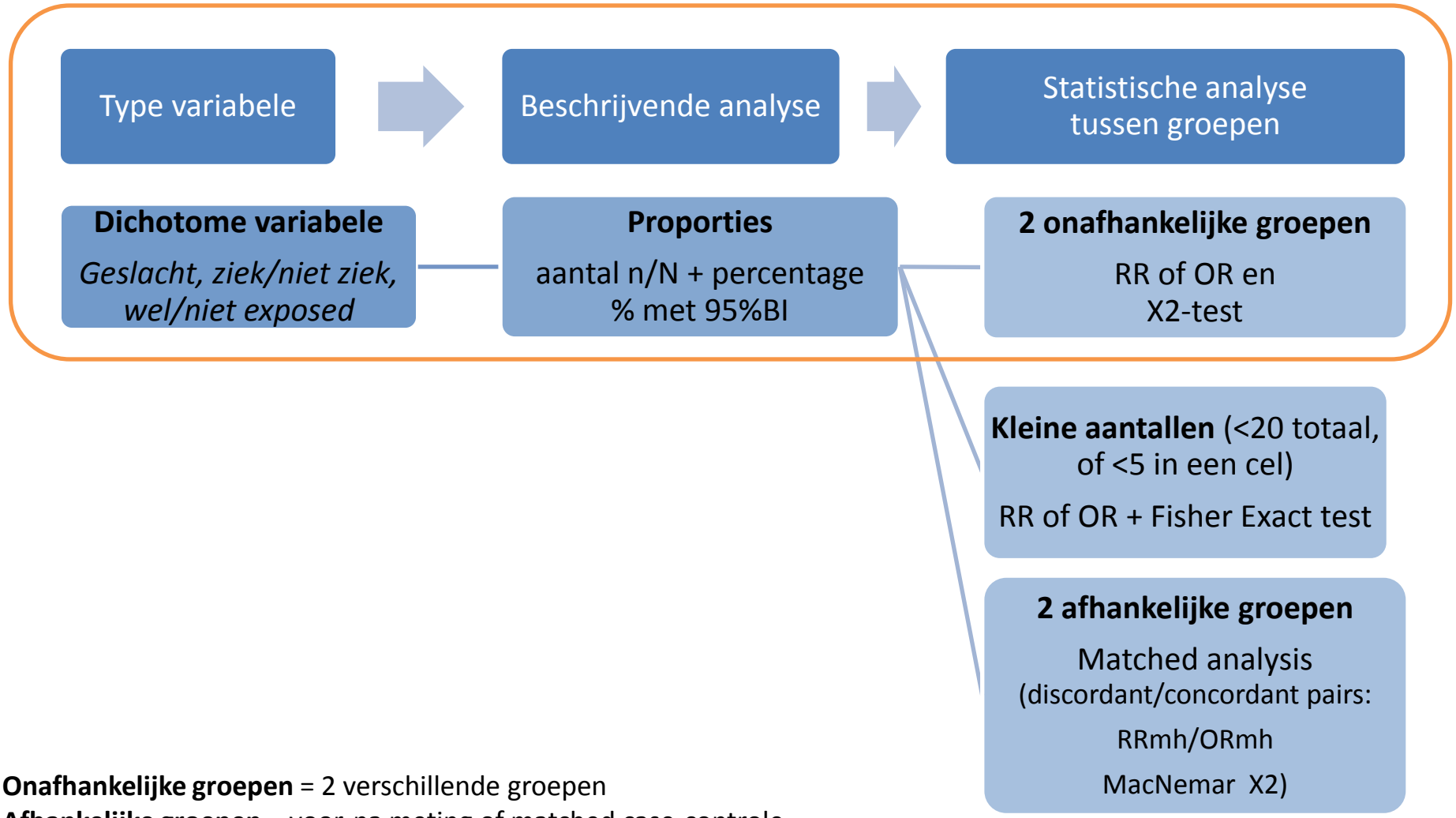
1. Vergelijk continue variabelen tussen 2 groepen



Onafhankelijke groepen = 2 verschillende groepen
Afhankelijke groepen = voor-na meting of matched case-controle

Beschrijvende en statistische analyses

2. Vergelijk dichotome variabelen tussen 2 groepen



Onafhankelijke groepen = 2 verschillende groepen
Afhankelijke groepen = voor-na meting of matched case-controle

Beschrijvende en statistische analyses

3. Vergelijk categorische variabelen tussen 2 groepen

Type variabele	Beschrijvende analyse	Statistische analyse tussen groepen
Categorisch variabele *≥2 categorieën leeftijdsgroep, opleidingsniveau*	**Proporties** aantal n/N + percentage % met 95%BI	**2 onafhankelijke groepen** RR/OR t.o.v. referentie categorie Categorisch: X2 (Ordinal: trend test)

* ANOVA geeft aan dat er een verschil is, maar niet waar het verschil zit.

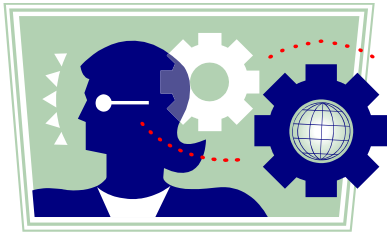
Beschrijvende en statistische analyses
4. Twee continue variabelen met elkaar vergelijken

Type variabele → Beschrijvende analyse → Statistische analyse **continu * continu**

Continue variabelen
Leeftijd, geslacht
CD4 en viral load

- **Normaal verdeeld**
  gemiddelde + SD
  → Pearson **correlatie coëfficient**
- **Niet normaal verdeeld (of lage aantallen: <20)**
  Mediaan (IQR)
  → Spearman Rho **correlatie coëfficient**

Interpretatie uitkomsten



- Kijk altijd EERST naar de beschrijvende gegevens: essentieel inzicht
- De effectmaat + betrouwbaarheidsinterval geven veel informatie
  - Omvang van het effect (sterkte associatie)
  - Richting van het effect (wordt iets groter of kleiner)
  - Precisie van de puntschatting van het effect (betrouwbaarheidsinterval)
- P-waarde is geen heilige graal
  - Statistische significantie is geen bewijs voor klinisch /biologisch effect
  - Afwezigheid van stat.significantie is geen bewijs voor geen effect – *"Absence of evidence is not evidence of absence"*
- Pas op: identieke P-waarden mogelijk bij:
  - Groot effect in een kleine populatie
  - Klein effect in grote populatie

Dus: kijk altijd naar de daadwerkelijke verschillen (of effectmaten)

- P-waarden en 95%BI zeggen niets over mogelijke bias/confounding

Voorbeeld varicella: Vraagstelling blijft centraal staan

- Veel variabelen zeer sterk geassocieerd met mening t.a.v. opname in RVP.
- (Analyse nog niet afgerond).
- **Maar**: niet alleen focussen op zeer sterke associaties met mooie P-waarden, de beschrijving is hier essentieel: 80% is 'tegen'

Bottom line: doe gedegen analysen en zorg dat het klopt, maar kijk altijd weer naar de onderzoeksvraag (hier: welk % is voor/tegen)

Variabelen, n/N (%)	Totaal (N=441)	RVP nee (N=350)	RVP ja (N=91)	Odds ratio	95% BI	p-value
Beroep, n/N (%)						
Verpleegkundige	279/441 (63.3)	233/350 (66.6)	46/91 (50.5)	Ref	-	-
Arts	162/441 (36.7)	117/350 (33.4)	45/91 (49.5)	1.95	1.22-3.11	0.005
Kinderen, n/N (%)						
Ja	338/440 (76.8)	270/350 (77.1)	68/90 (75.6)	Ref.	-	-
Nee	102/440 (23.3)	80/350 (22.9)	22/90 (24.4)	1.09	0.64-1.88	0.750
Waterpokken is een ziekte die je maar beter doorgemaakt kan hebben						
Niet mee eens	25/441 (5.7)	10/350 (2.9)	15/91 (16.5)	Ref.	-	-
Neutraal	77/441 (17.5)	60/350 (17.1)	17/91 (18.7)	0.19	0.07-0.50	0.001
Mee eens	339/441 (76.9)	280/35 (80.0)	59/91 (64.8)	0.14	0.06-0.33	<0.001

Voorbeeld TB in Tanzania: Duidelijke tabel en beschrijving resultaten

INT J TUBERC LUNG DIS 13(12):1524–1529
Vd Boogaard et al.

Tabel 2: Risicofactoren slechte behandeluitkomst in TBC patiënten in Tanzania in 2009 (N=2128)

	Behandeling succesvol N=2128	**Behandeling niet succesvol** N=641	**RR (95% BI)**	**P-waarde**
Overall, %	77%	23%		
Geslacht, n (%)				
Man	1343 (77%)	406 (23%)	Ref	-
Vrouw	785 (77%)	235 (23%)	1.0 (0.9-1.1)	0.97
HIV status, n (%)				
Negatief	753 (81%)	175 (19%) 928	Ref	-
Positief	626 (74%)	221 (26%) 847	1.4 (1.2-1.6)	<0.001
Onbekend	749 (75%)	245 (25%) 994	1.3 (1.1-1.6)	0.06
Type monitoring behandeling (DOT), n (%)				
Thuis	791 (70%)	332 (30%) 1123	1.7 (1.4-1.8)	<0.001
Kliniek	1337 (81%)	309 (19%) 1646	Ref	-

- Mensen met positieve HIV status hadden slechtere behandeluitkomst vergeleken met HIV negatieve TB patiënten (26% versus 19%, respectievelijk; RR 1.4, 95BI%...; P<0.0001).
- Patienten met thuis DOT hadden vaker slechtere behandeluitkomst (30%) vergeleken met kliniek-DOT (19%; RR 1.7; 95% CI 1.4-1.8; P<0.001)

1. Titel = volledig geeft weer **wat** er over **wie** in de tabel staat van **waar** en **hoeveel** (N)

2. Kolommen met de **twee groepen**, incl totale aantal (bijvoorbeeld N=2128)

3. Kolom met de **effectmaat** (dus het verschil of RR/OR) en de uitkomst **stat.test** (p-waarde) apart

12):1524–1529

Tabel 2: Risicofactoren slechte behandeluitkom TBC p

	Treatment succes N=2128	**Treatment unsuccesful** N=641		**RR (95% BI)**	**P-waarde**
Overall, %	77%	23%			
Geslacht, n (%)					
Man					-
Vrouw					0.97
HIV status, n (%)					
Negatief	753 (81%)	175 (19%)	928	Ref	-
Positief	626 (74%)	221 (26%)	847	1.4 (1.2-1.6)	<0.001
Onbekend	749 (75%)	245 (25%)	994	1.3 (1.1-1.6)	0.06
Type monitoring behandeling (DOT), n (%)					
Thuis	791 (70%)				
Kliniek	1337 (81%)		1646	0.6 (0.6-0.7)	<0.001

4. Per variabele weergegeven wat berekend is; in dit geval allemaal **n (%)**, maar kan ook **gemiddelde (SD)** zijn.

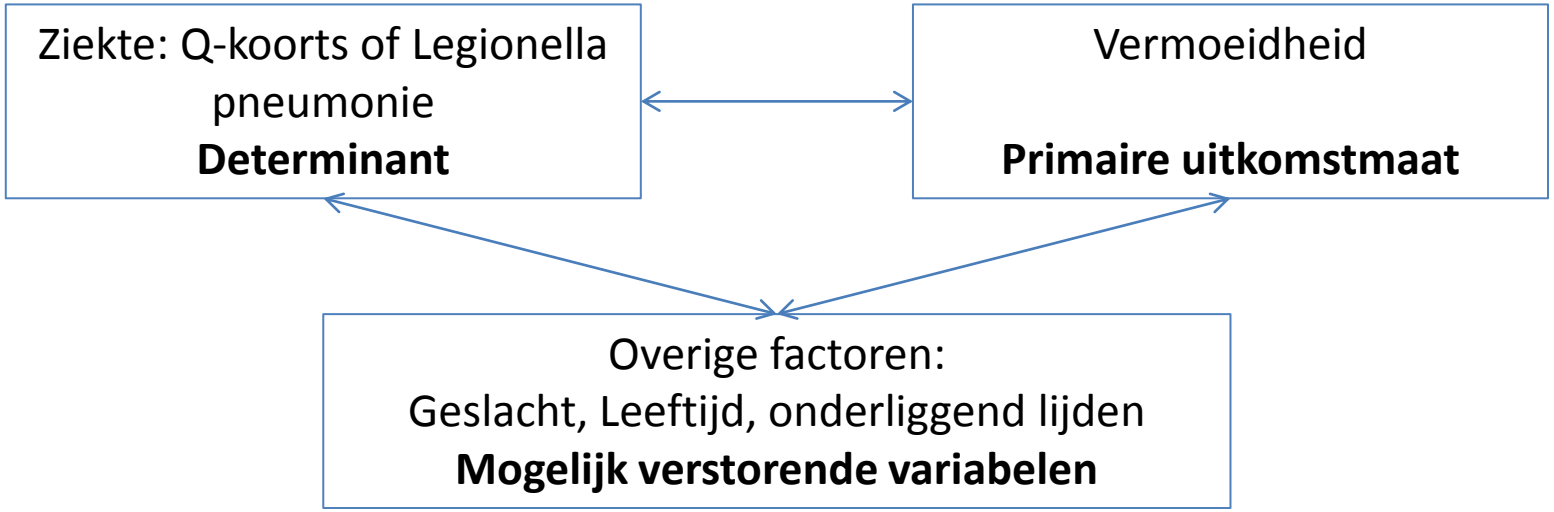
5. **verwoording van de resultaten**, inclusief begeleidende getallen en de effectmaat.

- Mensen met positieve HIV status hadden slechtere behandeluitkomst vergeleken met HIV negatieve TB patiënten (26% versus 19%, respectievelijk; RR 1.4, 95BI%...; P<0.0001).
- Patienten met thuis DOT hadden vaker slechtere behandeluitkomst (30%) vergeleken met kliniek-DOT (19%; RR 1.7; 95% CI 1.4-1.8; P<0.001)

Deel 3: Voorbeeld uit de praktijk

Joris van Loenhout – GGD VGGM (Arnhem) & AMPHI

- Onderzoek naar chronische vermoeidheid na Q-koorts
- Doelstelling onderzoek:
  - Vergelijken van vermoeidheidsstatus tussen patiënten met Q-koorts en Legionella pneumonie op 12 maanden na 1e ziektedag
- Dataverzameling: vragenlijsten
  - Instrumenten voor meten chronische vermoeidheid
  - Overige patiëntkarakteristieken



Voorbeeld dataverwerking II

	Q-koortspatiënten N=309	Patiënten met Legionella pneumonie N=190
Chronisch vermoeid, %	**60%**	**50%**
Mannelijk geslacht, %	54%	69%
Leeftijd (jaren), gemiddelde (SD)	50 (14)	61 (12)
Onderliggend lijden, %	41%	60%

- Legionella patiënten zijn vaker man, ouder en hebben meer onderliggend lijden
  - Alle drie risicofactoren voor chronische vermoeidheid
  - Alle drie relatie met zowel determinant als uitkomstmaat → dus factoren die je primaire onderzoeksvraag mogelijk verstoren…..
- Wat nu?
  - Ontdekken: is het confounding of effect modificatie? (Hulp statisticus)
  - Analyse uitbreiden: gecorrigeerde en gestratificeerde effectmaten berekenen om relatie Q-koorts & chronische vermoeidheid helder te krijgen.

Tip: altijd vragen opnemen in je vragenlijst om mogelijk verstorende variabelen te kunnen meten, zodat je er in de analysefase voor kunt corrigeren.

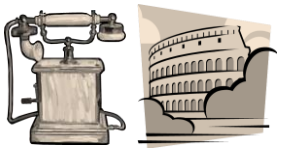
Take home messages

- Vraagstelling bepaalt onderzoeksopzet en de data-analyse
- Keuze statistiek bepaling bepaald door
  - Type variabele (continu – categorisch)
  - Type analyse (beschrijvend of vergelijkend)
  - Relatie tussen de groepen (onafhankelijk – afhankelijk)
- Raadpleeg epidemioloog/statisticus voor data-analyse (vóór start onderzoek)
- Bij twijfel over volgende stap in analyse: **terug naar onderzoeksvraag**

- Geef altijd **beschrijving van het effect** en de richting; nooit alleen P-waarde
- Statistische significantie ≠ bewijs voor klinisch /biologisch effect
- Klinische relevantie gaat boven statistische significantie: blijf dus zelf nadenken

- Als je de data-analyse fase zorgvuldig doorloopt dan zal je
  - Aan de hand van je verzamelde data antwoord kunnen geven op je onderzoeksvraag
  - Een goede basis hebben om conclusies te trekken en aanbevelingen te doen

Tot slot



- Vragen & suggesties
- Volgende Tele-ARENA: **donderdag 5 december** 2013, 12:00u – 13.00u
- Over een paar weken: evaluatie via digi vragenlijst (graag doorsturen naar Tele-ARENA deelnemers) – graag jullie respons!
  - Evaluatie deze serie:
    - Nut voor de GGD | Wat heb je geleerd? | Wat heb je gemist? | Opmerkingen
  - Toekomst:
    - Moet de Tele-ARENA doorgaan volgend jaar?
    - Verzoeknummers: welke onderwerpen moeten aan bod komen?
    - Andere suggesties en ideeën

Contact

Academische Werkplaats AMPHI

Radboud Universitair Medisch Centrum

Afdeling Eerstelijnsgeneeskunde

Geert Grooteplein Noord 21, Nijmegen

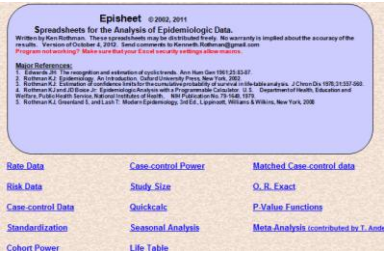
A.Tostmann@elg.umcn.nl – 024-3616632

Extra info bronnen

- Onderzoeksmethodologie:
  - Boek van *Kenneth Rothman: Modern Epidemiology*
- Sample size berekening –online tools:
  - http://www.stat.ubc.ca/~rollin/stats/ssize/
  - http://www.epibiostat.ucsf.edu/biostat/sampsize.html

Andere online tools voor o.a. sample size maar ook basic epi analyse

- Online EPI: **Open Source Epidemiologic Statistics for Public Health**
  - http://www.openepi.com/v37/Menu/OE_Menu.htm
- Epi-sheet
  - http://krothman.hostbyet2.com/Episheet.xls
- Epidemiology Net
  - http://www.epidemiolog.net/studymat/



Praktisch en zeer handzaam SPSS boek: Basisboek Statistiek met SPSS | D.B. Baarda, Cor van Dijkum & Martijn de Goede; 3e druk 2007.

Online vragenlijsten maken, versturen en data verzamelen: Lime Survey - www.LimeSurvey.org