

HC8 – testtheorie: signaal detectie theorie

Poppentest, (Koele, 1990), maken en nemen van beslissingen in onzekere omgeving. Bijv met slachtoffer van incest.

Betrek bij het beoordelen van de voorspellende gave (predictieve accuratesse) van een test het voorkomen van het kenmerk in de populatie (prevalentie)

Beslissend testgebruik

Signaal Detectie Theorie (SDT)

- Sensitiviteit
- Specificiteit
- Prevalentie
- Selectieratio
- Predictieve accuratesse

Grenswaarden

- Keuze grenswaarde (cut-off values)
- ROC curve (Receiver Operating Characteristic)
- Area Under Curve (AUC)

Psychologische diagnostiek:

- emotionele stoornis (depressie)
- cognitieve stoornis (dyslexie)
- gedragsproblemen (incest)
- geschiktheid (intelligentie)

Vergelijkbaar met medische diagnostiek (kanker, zwangerschap) en gebruik van detectieapparatuur (snelheidsmeter).

Beslissingen: aannemen of afwijzen; wel of geen behandeling

Centrale vraag: duiden waargenomen symptomen (signaal, testscore) op :

- noodzaak van behandeling?
- geschiktheid kandidaat?
- gevaar samenleving?

Beslissingen moeilijk:

- subjectieve oordelen van cliënt en diagnosticus.
- veelheid van informatie integreren tot een oordeel
- onzekerheid over testuitslag

Signaal detectie theorie (SDT)

theorie die mogelijkheid biedt dit soort vragen vanuit modelmatig, besliskundig perspectief aan te pakken. Denkkader en verzameling analysemethode. Theorie over kans op goede en foute beslissingen, met aannamen over verdeling van testuitkomsten onder verschillende omstandigheden.

Score hoger dan grenswaarde → stoornis

Lager dan grenswaarde → geen stoornis.

4 beslissingen:

		Gouden standaard		Total
		C+	C-	
Test	D+	a	b	a+b
	D-	c	d	c+d
Total		a+c	b+d	n

C+ behoort wel tot categorie	a = ware positieven (HITS)
C- behoort niet tot categorie	b = valse positieven (FALSE POSITIVES)
D+ positieve diagnose	c = valse negatieven (FALSE NEGATIVES)
D- negatieve diagnose	d = ware negatieven (CORRECT NEGATIVES)

Sensitiviteit = gevoeligheid van een kenmerk.

Hoe groot is de kans op een positieve diagnose gegeven dat iemand die stoornis heeft?

Kans op positieve diagnose (D+) gegeven dat het individu behoort tot categorie (C+).

$P(D+ | C+) = a/a+c \rightarrow$ voorwaardelijke of conditionele kans

Nb voor schatting sensitiviteit is een aselecte steekproef uit c+ nodig

Specificiteit: Hoe groot is de kans op een negatieve diagnose gegeven dat iemand die stoornis niet heeft?

Specificiteit: Kans op negatieve diagnose (d-) gegeven dat het individu behoort tot categorie (c-).

$P(D- | C-) = d/b+d$

Predictieve waarde

Positieve predictieve waarde (PPV): Proportie ware positieven gegeven een positieve diagnose.

$PPV = P(C+ | D+) = a/a+b$

Negatieve predictieve waarde (NPV): proportie ware negatieven gegeven een negatieve diagnose.

$NPV = P(C- | D-) = d/c+d$

Prevalentie (Base rate): Hoe groot is de kans dat iemand uit de populatie de stoornis heeft?

Kans op aanwezigheid stoornis (c+) voor iemand uit de populatie

$P(C+) = a+c/n$

Nb. voor schatting prevalentie is aselecte steekproef uit populatie nodig of informatie uit eerder onderzoek of door expertschatting.

Selectieratio: Hoe groot is de kans dat iemand uit de populatie een positieve diagnose (Ware en valse positieven) krijgt?

Selectieratio: ware en valse positieven samen. Kans op positieve diagnose (d+) voor iemand uit de populatie.

$P(D+) = a+b/n \rightarrow$ meestal niet goed.

NB. Voor schatting selectieratio is aselecte steekproef uit populatie nodig of informatie uit eerder onderzoek of door expertschatting.

$$P(D+) = P(D+ | C+) \cdot P(C+) + P(D+ | C-) \cdot P(C-)$$
$$= P(D+ | C+) \cdot P(C+) + (1 - P(D- | C-)) \cdot (1 - P(C+))$$

De kans op een positieve diagnose ongeacht aanwezigheid stoornis = sensitiviteit x prevalentie + (1 - specificiteit) x (1 - prevalentie)

Regel van Bayes voor predictieve accuratesse:

$$P(C+ | D+) = P(D+ | C+) \cdot \frac{P(C+)}{P(D+)} = \text{sensitiviteit} \times \frac{\text{prevalentie}}{\text{selectieratio}}$$

Predictieve accuratesse

Predictieve accuratesse van diagnostische test is afhankelijk van:

- kwaliteit instrument (sensitiviteit en specificiteit)
- voorkomen van stoornis, ziekte, trauma (prevalentie) \rightarrow Zelfs bij betrouwbaar instrument is bij lage prevalentie de kans op correcte beoordeling klein.

Base rate neglect: Niet betrekken van (lage) prevalentie bij beoordeling van cliënt op basis van positieve diagnose en de sensitiviteit van de test.

- predictieve accuratesse is bij lage prevalentie altijd gering
- met subjectieve apriori verwachting kun je $P(C+)$ aanpassen

Grenswaarde β : Hoe goed kan je met een test de diagnose 'dit is een pester' stellen?

Algemeen: wat is bij diagnostiek met test scores een verstandige keuze voor grenswaarde β (**cut off score**)

Probleem: Hoe kiezen we op verstandige wijze (rationeel) de grenswaarde β ?

Proportie ware positieven neemt af bij strengere grenswaarde (meer naar recht). Proportie ware negatieven neemt toe bij strengere grenswaarde (meer naar links).

1. Optimale grenswaarde

Subjectieve nut/waarde vaststellen van ware positieven, valse positieven, ware negatieven en valse negatieven

2. Subjectieve nut

Hoe belangrijk is correcte beslissing?: ware positieve; ware negatieve

Hoe erg is foute beslissing?: valse positieve; valse negatieve
Keuze grenswaarde ? geen pester wel pester

→ opbrengstmatrix

Gewenst:

- Hoog percentage *ware positieven*

Alleen die leerlingen te pakken krijgen die echte pesters zijn → hoog percentage *ware positieven* (= hoge sensitiviteit)

- Laag percentage *valse positieven*

Geen leerlingen als pester aanwijzen terwijl ze het niet zijn. → laag percentage *valse positieven* (= lage (1 – specificiteit))

Hulpmiddel

Bekijk voor alle grenswaarden de sensitiviteit en de (1 – specificiteit): → ROC-curve

SPSS: Menu; Analyze; ROC-curves

Receiver operating characteristic (ROC) – curve: Voor een test zet je voor alle mogelijke grenswaarden tegen elkaar af:

- proportie ware positieven (Sensitiviteit)
- proportie valse positieven (1-specificiteit)

Snijpunt diagonaal met curve is positie met beste grenswaarde.

Area under curve: Maat voor testkwaliteit. Gevoeligheid van test voor onderscheiden van diagnostische categorieën (= testsensitiviteit) onafhankelijk van gekozen grenswaarde. Alles onder de curve!

Minimale sensitiviteit: AUC = .50; Maximale sensitiviteit: AUC = 1.0; redelijke sensitiviteit: AUC = .75

Grenswaarde voor beoordeling schaal

Oppervlak onder de curve (AUC)

> .90 uitmuntend

.80 - .90 goed

.70 - .80 redelijk

.60 - .70 matig

.50 - .60 slecht

HC9 – diagnostiek: klinische psychologie

Wat valt er onder? Psychiatrische stoornissen en gerelateerde klachten.

DSM-5:

- Laag gewicht (BMI)
- Verstoord (restrictief) eten
- Verstoord lichaamsbeeld
- Comorbiditeit (hoge angst / depressie)
- Neuropsychologische problemen
- Fysieke / somatische problemen

Is er sprake van anorexia nervosa? Brief verwijzer → eetstoornis gerelateerde pathologie → anamnese → heteroanamnese → somatisch onderzoek

Eating disorder examination: Fairburn.

4 subschalen: lijnen, piekeren over eten, piekeren over lichaamsvormen, piekeren over gewicht.

Anamnese: uitzoeken wat de patiënt wil, wat is het probleem?

Heteroanamnese: ouders, partner, vriend(in), kinderen. Wat zien zij?

Somatisch onderzoek: uitblijven menstruatie, temperatuur, vertraagde bloedsomloop, bloedwaarden, vermoeidheid e.d.

Cognitieve rigiditeit: de manier waarop je denkt. Wanneer je cognitief rigide bent krijg je een soort 'kortsluiting'

- Set-shifting
- Centrale coherentie.

Planning kun je testen met tower of London Test.

Rey-osterrieth complex figure: opdracht om een figuur na te tekenen.

- Centrale coherentie: seeing the bigger picture (versus detail focus)

Opvallend aan instrumenten is dat persoonlijkheid vaak gemeten wordt met self-report vragenlijsten (subjectief) en executief functioneren en neuropsychologische functies worden vaak getest met taken.

HC10 – diagnostiek: bias in de psychologische diagnostiek

Een testitem is niet zomaar wat het lijkt, verwachtingen over de testsituatie kunnen de testuitslag bepalen (bijv. moeilijker denken dan het is).

Bias: systematische fout in een testuitslag (Reynolds & Livingston)

Vertekening die de validiteit van een testuitslag bedreigt (van de Vijver, 2011)

CTB hypothese: groepsverschillen te wijten aan een inherente fout in de test

Perspectieven hoe je tegen bias aan kan kijken:

- Statistisch
- Juridisch: jaren '60 – jaren '80, zijn testuitslagen valide genoeg om ze in een rechtzaak te gebruiken?
- Filosofisch / ethisch: wat betekent bias ethisch gezien? Mogen we wel op grond van de test beslissingen nemen?
- Sociaal - cultureel: CITO-scores, mogen we deze openbaar maken? E.d.
- Klinisch / praktisch: als je weet dat een test mogelijk bias heeft voor een groep, wat betekent dit in de praktijk? Welke waarde moet je daar aan hechten?

Fairness: onpartijdige toepassing van test:

- Zonder bias
- Gelijke testomstandigheden
- Gelijke gelegenheid ter voorbereiding / leren
- Gelijke groepsuitkomsten (voor zo ver mogelijk)

Bias = unfair

Unbiased ≠ fair

3 vormen van vertekening bij een test:

- constructbias: meet niet zoals het zou horen te meten
- methodebias: afname, afnemer, interactie testafnemer en geteste
- itembias: een specifiek item in een test beoordeelt sommige groepen anders.

H&M beweren (VS): intelligentie / iq voorspelt maatschappelijke variabelen zoals inkomen en werk, etnische verschillen in IQ tussen blanke en zwarte Amerikanen is gemiddeld 15 punten (SD = 1).

Gould: er wordt uitgegaan van 4 verschillende aannames:

1. intelligentie is een getal
2. dat getal kun je in rang ordenen (en correleert met maatschappelijke verschillen)
3. intelligentie is erfelijk
4. intelligentie is onveranderlijk

Huidige situatie in de VS: in 1999 hebben APA en AERA Standards opgesteld: fairness is het belangrijkste aandachtspunt bij toetsen en het beoordelen van de kwaliteit van een test. Ook wordt er meer aandacht besteedt aan fairness dan aan bias en er is bijzondere aandacht voor fairness bij minderheden, meertalige kinderen en mensen met beperkingen.

Achtergronden in NL:

- weinig interesse ('60 en '70)
- W.B. Hofstee signaleert en analyseert het probleem (eind '80)
- Onderzoek naar bias in IQ en assessment tests ('90)
- Ontwikkeling van nieuwe tests, bruikbaar in multiculturele samenleving (vanaf '90)

Hofstee e.a. hebben de toepasbaarheid van psychologische tests bij alloctonen onderzocht (1990): 20 tests, ethnocentrisme in bijna alle tests (bias dominante cultuur) en er waren geen normgegevens beschikbaar over de toepasbaarheid van de test voor alloctonen. De tests waren dus beperkt bruikbaar.

Als gevolg van dit onderzoek heeft Hofstee de COTAN gestimuleerd om het criterium bias en toepasbaarheid op te nemen, hij vond dat er onderzoek nodig was naar testbias in NL en herziening van subtests in verband met ethnocentrisme. Ook was er grote terughoudendheid bij toepassing van tests bij allochtonen (i.h.b. bij werving en selectie)

Huidige situatie in NL: Bochhah e.a.

- Meer onderzoek, nieuwe tests ontwikkeld
- Aan bestaande tests is niks veranderd
- Boek: deskundigen over testen van etnische minderheden
- 2005: richtlijnen voor het gebruik van diagnostische instrumenten bij etnische minderheden.
 - Bijna niemand kent / gebruikt deze richtlijnen (worden ook niet gebruikt bij opleidingen)
- Huijding e.a. (2012): voorzet voor verantwoord en fair testgebruik en de rol van de COTAN

Oplossingen of schijnoplossingen?

- Cultuurvrije tests: taal, spreekwoorden, abstracte begrippen
 - Niet mogelijk
- Standaardtests: voorspellende waarde IQ- en capaciteitentests is goed, maar niet alle tests beogen te voorspellen.
- Geen testgebruik, maar interviews en observaties: nog minder validiteit, kind weg met badwater (en-en ipv of-of)

Nu: kunnen gestandaardiseerde test verantwoord en fair worden gebruikt bij diverse groepen?

Drie domeinen van testgebruik: arbeidsmarkt (werving en selectie), onderwijs (diagnostiek, behandeling en verwijzing) en geestelijke gezondheidszorg (intake, diagnostiek, monitoring effecten)

Verantwoord testgebruik (voor de afname): controleer de testbaarheid van de deelnemer, mocht Nederlands niet de eerste taal zijn neem dan de test die zo min mogelijk beroep doet op de taalvaardigheden. Gebruik als dat mogelijk is altijd meer dan één instrument, controleer de validiteit en ga de groepsverschillen na. Controleer ook begrippen, symbolen, afbeeldingen en check mogelijke aanpassingen die je zou mogen of kunnen doen.

Verantwoord testgebruik (tijdens de afname): informeer over het doel van de test, de procedure en de tijdsduur.

Verantwoord testgebruik (resultaten en interpretatie): normgroepen, bij twijfel over de validiteit geen conclusies trekken. Wat is de waarde van de testscore in de eindbeoordeling en wat is het leerpotentieel? Leg veranderingen en overwegingen vast.

Voorbeelden zie collegeslides.

HC11: Psychologische diagnostiek praktijkveld neuropsychologie

Waarom neuropsychologie apart? Het is een wezenlijk andere populatie (somatisch), multidisciplinair en ander onderzoek (psychologisch test onderzoek, gedrag uitlokken).

Neuropsychologie: cognitie, emotie en gedrag.

Kijken vooral naar hoe het op dit moment gaat, zijn er nu cognitieve problemen?

Anamnese:

- Basis principe gelijk aan KP
- Patiënt ipv cliënt (lopen vaak ook al bij een geriater, e.d.)
- 'reason to encounter' (je komt niet zomaar, de kans dat je iets vindt is groter)
- Aanvrager is veelal een arts (je werkt parallel, multidisciplinair)
- Angst (voor slecht maar ook voor goed nieuws, bijv. bij werk)
- (semi-)gestructureerd
- Anders dan een 'intake'

Wat is voor hen de belangrijkste rede?

Niet alleen de klacht maar ook de verwachting wordt daarmee uitgesproken

Inzicht in eigen functioneren?

Ziektebesef vs. Ziekte inzicht

Taal keuze en taal beperkingen: open en gesloten vragen, sturend en niet sturend.

Heteroanamnese:

Zeer belangrijk: inzicht, streven naar volledigheid, moeilijke problematiek, apart óf samen (loyaliteitsconflicten), betrouwbaarheid.

Testonderzoek:

- Hypotheses?
- Belastbaarheid?
- Observatie (vb begripsniveau, angst, frustatietolerantie)
- Alle cognitieve domeinen
 - Wel met richting en voorkomen van redundantie
- Streven naar volledigheid op een efficiënte, betrouwbare en valide wijze
- In staat om bij te stellen onderweg
- Accommodations Adaptive Testing

De score is niet alleen de score, maar ook de weg naar de score toe → observeren

Het is een momentopname !

2 of meer standaarddeviaties onder het gemiddelde: stoornisniveau.

Testen zijn 'slechts' een middel om gedrag uit te lokken, hypothesen zijn daarom noodzakelijk, plus kennis van cognitieve concepten en een goede observatie.

Interpreteren:

- Wat zie je? Blijf bij de resultaten
- Eerst beschrijven wat je hebt
- Wat waren de hypothesen?
- Soms weer op nieuw beginnen of een stukje terug gaan
- Komt het overeen met je indruk?
- Kun je het verdedigen? Zo nee: verander het → terug naar resultaten

Rapporteren: aan iedereen dezelfde boodschap, universele taal, ondubbelzinnig, het moet een antwoord zijn of een stap verder in het diagnostische proces.

Rapporteren aan de patiënt:

- Bij BIG: als eerste door de neuropsycholoog
- Veiligheid van de patiënt
- Houd rekening met wat je weet (aandachts- of taalproblemen e.d.)
- Word je nog gevolgd of loop je alleen?
- Is het herkenbaar voor de patiënt?
- Komt het overeen met de verwachting?

Rapporteren aan de aanvrager/team:

Ethiek: een tak van de filosofie die zich bezighoudt met de kritische

Wanneer kom je met ethiek in aanraking? Altijd. Je bent onderdeel van de maatschappij.