

gen af UL, tydning af UL-billeder og de specifikke teknikker ved de enkelte blokader. Dertil spiller udstyr og supervision en stor rolle.

Målsætningen for de næste år i pædiatrisk regionalanæstesi synes at være mere forskning på området og øget anæstesiologisk ekspertise i brugen af UL.

KORRESPONDANCE: Michèle Reeslev Lefort, Anæstesi- og Operationsafsnit 4013, Juliane Marie Centret, Rigshospitalet, 2100 København Ø.
E-mail: mlefort@gmail.com

ANTAGET: 24. november 2010

FØRST PÅ NETTET: 4. juli 2011

INTERESSEKONFLIKTER: ingen

LITTERATUR

1. Willschke H, Marhofer P, Machata A-M et al. Current trends in paediatric regional anaesthesia. *Anaesth* 2010;65(suppl 1):97-104.
2. Willschke H, Marhofer P, Bösenberg A et al. Epidural catheter placement in children: comparing a novel approach using ultrasound guidance and a standard loss of resistance technique. *Br J Anaesth* 2006;97:200-7.
3. Tsui BCH, Suresh S. Ultrasound imaging for regional anaesthesia in infants, children, and adolescents. *Anesth* 2010;112:719-28.
4. Marhofer P, Bösenberg A, Sitzwohl C et al. Pilot study of neuraxial imaging by ultrasound in infants and children. *Paediatr Anesth* 2005;15:671-6.
5. Willschke H, Marhofer P, Bösenberg A et al. Ultrasonography for ilioinguinal/iliohypogastric nerve blocks in children. *Br J Anaesth* 2005;95:226-30.
6. Marhofer P, Sitzwohl C, Greher M et al. Ultrasound guidance for infraclavicular brachial plexus in children. *Anaesthesia* 2004;59:642-6.
7. Oberdorfer U, Marhofer P, Bösenberg A et al. Ultrasonographic guidance for sciatic and femoral nerve blocks in children. *Br J Anaesth* 2007;98:797-801.
8. Koscielniak-Nielsen ZJ. Ultrasound-guided peripheral nerve blocks: what are the benefits? *Acta Anaesthesiol Scand* 2008;52:727-37.
9. Tsui BCH, Suresh S. Ultrasound imaging for regional anaesthesia in infants, children, and adolescents. *Anesth* 2010;112:473-92.
10. Rozen WM, Tran TM, Ashton MW et al. Refining the course of thoracolumbar nerves: a new understanding of the innervation of the anterior abdominal wall. *Clin Anat* 2008;21:225-33.
11. Suresh S, Chan VW. Ultrasound guided transversus abdominis plane block in infants, children and adolescents: a simple procedural guidance for their performance. *Paediatr Anesth* 2009;19:296-9.
12. Petersen PL, Mathiesen O, Torup H et al. The transversus abdominis plane block: a valuable option for postoperative analgesia? *Acta Anaesth Scand* 2010;54:529-35.
13. Ivani G, Mosetti V. Pediatric regional anaesthesia. *Minerva Anestesiol* 2009;75:577-83.
14. van Schoor AN, Boon JM, Bösenberg A et al. Anatomical considerations of the pediatric ilioinguinal/iliohypogastric nerve block. *Paediatr Anaesth* 2005;15:371-7.
15. Weintraud M, Marhofer P, Bösenberg A et al. Ilioinguinal/iliohypogastric blocks in children: where do we administer the local anesthetic without direct visualization? *Int Anesth Res Soc* 2008;106:89-93.
16. Courrges P, Poddevin F. Rectus sheath block in infants: what suitability? *Paediatr Anesth* 1998;8:181-2.
17. Ferguson S, Thomas V, Lewis I. The rectus sheath block in pediatric anaesthesia: new indications for an old technique? *Paediatr Anesth* 1996;6:463-6.
18. Willschke H, Bösenberg A, Marhofer P et al. Ultrasonography-guided rectus sheath block in paediatric anaesthesia – a new approach to an old technique. *Br J Anaesth* 2006;97:244-9.
19. Motoyama EK, Davis PJ. *Smith's anesthesia for infants and children*. 7th ed. Philadelphia: Mosby Elsevier, 2006.
20. Lönnqvist PA, Olsson GL. Paravertebral vs epidural block in children. *Acta Anaesthesiol* 1994;38:346-9.
21. Eng J, Sabanathan S. Continuous paravertebral block for post-thoracotomy analgesia in children. *J Pediatr Surg* 1992;26:556-7.
22. Berta E, Spanhel J, Smakal O et al. Single injection paravertebral block for renal surgery in children. *Paediatr Anesth* 2008;18:593-7.
23. Splinter WM, Thomson ME. Somatic paravertebral block decreases opioid requirements in children undergoing appendectomy. *Can J Anesth* 2010;57:206-10.
24. Naja Z, Raf M, El Rajab M et al. A comparison of nerve stimulator guided paravertebral block and ilio-inguinal nerve block for analgesia after inguinal herniorrhaphy in children. *Anaesthesia* 2006;61:1064-8.
25. Renes SH, Bruhn J, Gielen MJ et al. In-plane ultrasound-guided thoracic paravertebral block. *Reg Anesth Pain Med* 2010;35:212-6.
26. O'Riain SC, Donnell BO, Harmon DC et al. Thoracic paravertebral block using real-time ultrasound guidance. *Anesth Analg* 2010;110:248-51.
27. Vaghadia H, Jenkins LC. Use of a Doppler ultrasound stethoscope for intercostal nerve block. *J Can Anaesth* 1988;35:86-9.
28. Byas-Smith MG, Gulati A. Ultrasound-guided intercostal nerve cryoablation. *Anesth Analg* 2006;103:1033-5.
29. Markakis DA. Regional anaesthesia in pediatrics. *Anesthesiol Clin North America* 2000;18:355-81.
30. Nakagawa M, Hayashi M, Houki S et al. Ultrasound guided T2 intercostal nerve block: a comparison with stellate ganglion block under the blind technique and ultrasound guided technique. *Masui* 2010;59:604-9.

Forskellig betydning af p-værdien i eksplorerende og konfirmerende hypotesetestning

Oke Gerke¹, Poul Flemming Højilund-Carlsen¹ & Werner Vach²

Evidensbaseret medicin opfordrer lægerne til at tage kliniske beslutninger ud fra eksisterende viden, gældende retningslinjer og diverse referenceprogrammer. *State of the art* fremgår af nyligt offentliggjorte *peer review*'ede artikler og metaanalyser af disse. Klinisk videnskab kan karakteriseres ved fire forskellige faser i lighed med dem, som kendes fra lægemiddelindustrien. Det er især en god ide at skelne imellem eksplorerende (fase II) og konfirmerende (fase III) studier. Konklusioner, der er draget på baggrund af konfirmerende studier, har en helt anden vægt end konklusioner, der stammer fra eksplorerende undersøgelser. Ikke desto mindre er en p-værdi under 5%

grænsen blevet et mantra for den kliniske videnskab: Det ultimative mål er at opnå (statistisk) signifikans uden skelen til den enkelte undersøgelses nærmere karakter. I denne artikel diskuteres betydningen og anvendelsen af statistisk hypotesetestning og i særdeleshed fortolkningen af p-værdien i et forsøg på at kaste lys på konceptuelle forskelle imellem eksplorerende og konfirmerende hypotesetestning, som har hver deres formål.

DET VIDENSKABELIGE SPØRGSMÅL

Sædvanligvis omfatter sund videnskab en kronologisk række af begivenheder: 1) formulering af et viden-

STATUSARTIKEL

1) Nuklearmedicinsk Afdeling, Odense Universitetshospital, og
2) Institut für Medizinische Biometrie und Medizinische Informatik, Albert-Ludwigs-Universität Freiburg, Tyskland



TABEL 1

Mulige udfald af hypotesetestning.

Beslutning på baggrund af eksperiment til fordel for	Sandhed	
	nulhypotese	alternativ hypotese
Nulhypotese	Korrekt beslutning	Type II-fejl
Alternativ hypotese	Type I-fejl	Korrekt beslutning

skabeligt spørgsmål og opstilling af forsøgsdesign, 2) indsamling af data, der kan besvare spørgsmålet, og 3) analyse af resultater og formidling.

Kun præcise spørgsmål muliggør en strikt kronologi. I modsætning hertil fordrer vage ideer en efterfølgende proces, som består i at finde og formulere et specifikt og relevant videnskabeligt spørgsmål, som rent faktisk kan besvares ud fra de data, som skal indsamles [1]: 1) ide, 2) iterativt: spørgsmål, design, spørgsmål, design etc., 3) dataindsamling, 4) analyse og 5) besvarelse.

HYPOTESETESTNING, TYPE I- OG TYPE II-FEJL, STYRKE OG P-VÆRDI

Da direkte verifikation af hypoteser er umulig, baseres hypotesetestning på princippet om falsifikation [2]. Hertil kræves to hypoteser: en nulhypotese og en alternativ hypotese. Sidstnævnte er den originale forskningshypotese, mens den komplementære nulhypotese repræsenterer »nul difference«. Hvis man f.eks. sammenligner ny og gammel behandling, indikerer nulhypotesen »ingen forskel i behandlingseffekt«, hvorimod den alternative hypotese lyder: »Der er en forskel i behandlingseffekt«. En af hypoteserne er sand, men vi ved ikke hvilken. Resultatet af forsøget kalder på en beslutning til fordel for en af hypoteserne. Der er fire mulige udfald: to beslutninger er korrekte, to er ukorrekte. De repræsenterer henholdsvis type I- og type II-fejl (Tabel 1).

En type I-fejl foreligger, når nulhypotesen »ingen forskel« forkastes, selv om den er sand. En type II-fejl betyder, at man accepterer nulhypotesen til trods for, at der er en sand forskel. Det forhold, at en beslut-

ning til fordel for en falsk behandlingsforskel (type I-fejl) potentielt er mere farlig for patienten, end hvis man holder sig til det etablerede behandlingsregime, selv om der findes et bedre alternativ (type II-fejl), bevirker, at sandsynligheden for at begå en type I-fejl i planlægningsfasen sædvanligvis sættes væsentligt lavere end sandsynligheden for at begå en type II-fejl, f.eks. ved henholdsvis 5% og 20%. Type II-fejlen har betydning for studiets såkaldte styrke (styrke = 1 – sandsynligheden for type II-fejl), hvilket er sandsynligheden for at forkaste nulhypotesen til fordel for en sand alternativ hypotese før eksperimentet. Styrken afhænger af stikprøvens størrelse, baseres på test af en enkelt hypotese og sættes sædvanligvis til 80%. p-værdien er sandsynligheden for at opnå de resultater, som ses i undersøgelsen (eller endnu mere ekstreme værdier til fordel for den alternative hypotese), hvis nulhypotesen er sand. Hvis p-værdien er under signifikansgrænsen $\alpha = 0,05$ (eller 5%), konkluderes det, at der er statistisk signifikans. p-værdien siger intet om, hvor sandsynlig nulhypotesen er [3]. Hvad mere er, den fortæller ikke noget om størrelsen af den pågældende behandlingseffekt, ej heller noget om dens kliniske værdi. Derfor er brugen af ledsagende konfidensintervaller altid på sin plads [4].

MULTIPLE TEST OG DEN EKSPERIMENTRELATEREDE TYPE I-FEJL

Niveauet for en type I-fejl refererer kun til en enkelt hypotesetest. Så snart der udføres multiple parallelle test på det samme datamateriale, stiger sandsynligheden for, at der begås en type I-fejl i det samlede antal test, dvs. sandsynligheden for at forkaste mindst én nulhypotese, selv om den er sand, så resultatet bliver en falsk positiv beslutning. Hvis der f.eks. udføres fem statistiske test af fem sideordnede hypoteser, stiger signifikansniveauet for hele eksperimentet til $1 - (0,95)^5 = 0,23$ (23%), og for ti hypoteser til $1 - (0,95)^{10} = 0,40$ (40%). Fem eller ti hypoteser kan lyde af mange, men er typiske antal. Hvis der ikke korrigeres for massesignifikans, opnås en række falsk positive resultater på 5%-niveauet, hvilket indebærer, at p-værdier under dette niveau ikke kan opfattes som alvorlige indikatorer imod nulhypotesen.

Der er imidlertid måder, hvorpå man kan kontrollere den slags eksperimentrelaterede fejl, heriblandt Bonferroni-Holm-metoden [5], som er en af de mest benyttede post hoc-procedurer, når endepunkterne er uafhængige. Ved brug af denne skal p-værdierne sammenlignes med hensyn til niveauer under de 5%, for at man kan holde den eksperimentrelaterede type I-fejl nede på 5%. Er hypoteserne afhængige, f.eks. fordi der testes for effektivitet af forskellige doser, kan denne korrelation udnyttes for at holde denne fejl nede på



FAKTABOKS

Statistisk signifikante resultater ($p < 0,05$) øger chancen for publicering.

Ureflekteret generering af flere p-værdier kan øge den eksperimentrelaterede type I-fejl (og dermed risikoen for falsk positive konklusioner).

Den primære hypotese testes konfirmerende, sekundære hypoteser testes kun eksplorerende.

Konfirmerende testning for flere primære endepunkter kan kræve korrektion for multipel testning og dermed et større patientantal.

p-værdien skal angives med to ($p \geq 0,01$) eller tre decimaler ($p < 0,01$) og ikke vurderes dikotomiseret ($<$ eller $> 0,05$) og isoleret, men i sammenhæng med effektstørrelsen, dens konfidensinterval, potentielle konfoundere og evidens fra andre studier.

5%, idet der testes faldende doser i en lukket testprocedure (hierarkisk testning).

EKSPLORERENDE VERSUS KONFIRMERENDE TESTNING ELLER: DET PRIMÆRE ENDEPUNKT – ET VANSKELIGT VALG

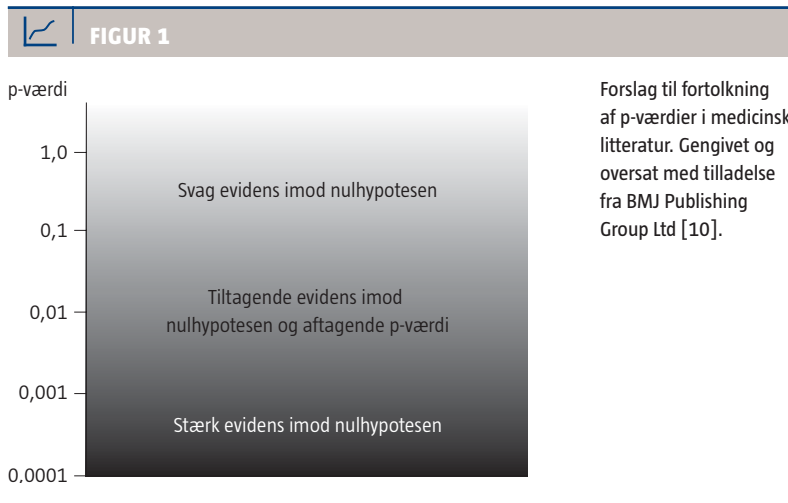
For at undgå fejl som følge af massesignifikans udvælges der ofte en enkelt hypotese, som testes i henhold til de ovenfor beskrevne strikte regler for en konfirmerende undersøgelse, mens de resterende hypoteser angives kun at have eksplorerende karakter. Denne fremgangsmåde realiseres som regel ved på forhånd i projektplanen at definere ét primært endepunkt og et eller flere sekundære endepunkter, som gøres til genstand for henholdsvis konfirmerende og eksplorerende testning. Det er således kun den primære hypotese, som testes konfirmerende på et givet signifikansniveau, og som danner basis for vurdering af den nødvendige stikprøvestørrelse. Den konfirmerende test bør have en tilstrækkelig høj styrke til, at den primære hypotese kan af- eller bekræftes. De øvrige hypoteser testes ofte uden korrektion for massesignifikans. Eksplorerende hypoteser kan stamme fra litteraturen og blive testet for at se, om de også holder i det aktuelle studium (f.eks. i en bestemt undergruppe). Mere almindeligt er dog, at de eksplorerende hypoteser også er »nye«, og at det er nødvendigt at bekræfte dem i et efterfølgende nyt eksperiment, hvori den eksplorerende hypotese bliver den konfirmerende hypotese.

FALDGRUBER I LITTERATUREN

Uheldigvis kan man ikke altid i de såkaldte *peer-review*'ede artikler læse sig til, om det anførte primære endepunkt også var det, man oprindeligt havde planlagt. Selv om en undersøgelse blev registreret, inden den blev udført (f.eks. [6]) kan det samme materiale vise sig at være blevet analyseret adskillige gange med hensyn til en række spørgsmål, som hver har ført til en selvstændig publikation. Dette er i sig selv også en form for gentagen testning uden korrektion for massesignifikans. Når forfatterrækkefølgen, som det ikke sjældent er tilfældet, også er ændret, kan det være ganske vanskeligt at opdage [7]. Mens en enkelt ud af en serie artikler, der er baseret på det samme materiale, kunne se ud til at repræsentere et konfirmerende studie, foreligger der rent faktisk kun en rapport af eksplorerende karakter, idet den eksperimentrelaterede fejl er pustet op til et niveau højt over den konventionelle grænse for en type I-fejl.

DEN MAGISKE P-VÆRDI

p-værdien bevirker dikotomisk inddeling af resultater som værende enten statistisk signifikante eller stati-



stisk ikkesignifikante. Den rummer ingen information om f.eks. størrelsen af en observeret behandlingseffekt, og fundet af $p < 0,05$ kan lige så vel stamme fra et forsøg med en meget lille behandlingsforskel i et studie med mange patienter som fra en undersøgelse med en stor terapeutisk forskel observeret i et ganske lille materiale [3]. p-værdiens betydning kan således let overvurderes, hvis den benyttes som det eneste udtryk for succes.

I 1994 satte *Altman* spørgsmålstegn ved kvaliteten af klinisk forskning, fordi den tilsyneladende drejede sig mere om kvantitet end kvalitet [8]. Misbrug af statistiske test til at generere statistisk signifikante resultater blev eksemplificeret af *James Mills* i 1993 [9]. Han beskrev to former for datatortur, hvoraf den ene omfattede multiple test og massesignifikans. Der er siden fremsat forslag om en mere adækvat håndtering af p-værdier, herunder angivelse af de faktiske p-værdier i stedet for oplysning om, at p-værdien lå under 5%-niveauet, fortolkning af de medicinske resultater i lyset af studietype og anden tilgængelig evidens og endelig overvejelser om bias eller *confounding factors* i de tilfælde, hvor der foreligger lave p-værdier [10]. I stedet for dikotomisering af resultater til enten at være statistisk signifikante eller statistisk ikkesignifikante, bør p-værdien fortolkes i et kontinuerligt spektrum (**Figur 1**).

KONKLUSION

Hypotesetest skal anvendes rigtigt med a priori-definerede endepunkter, relevant korrektion for eventuel massesignifikans samt forsigtig og nuanceret fortolkning af de resulterende p-værdier. Effektstørrelse og ledsagende konfidensintervaller bør altid angives. p-værdier fra eksplorerende hypotesetestning skal opfattes eksplorerende, muligvis hypotesegenererende, dog ikke konfirmerende.

KORRESPONDANCE: Oke Gerke, Nuklearmedicinsk Afdeling, Odense Universitets-hospital, Sdr. Boulevard 29, 5000 Odense C.
E-mail: oke.gerke@ouh.regionsyddanmark.dk

ANTAGET: 15. april 2011

FØRST PÅ NETTET: 16. maj 2011

INTERESSEKONFLIKTER: ingen

LITTERATUR

1. Tukey JW. We need both exploratory and confirmatory. *Am Stat* 1980;34:23-5.
2. Popper K. The logic of scientific discovery. New York: Harper & Row, 1934/1959.
3. Goodman SN. Toward evidence-based medicine. 1: The p value fallacy. *Ann Intern Med* 1999;130:995-1004.

4. Altman DG. Why we need confidence intervals. *World J Surg* 2005;29:554-6.
5. Holm S. A simple sequentially rejective multiple test procedure. *Scand Stat Theory Appl* 1979;6:65-70.
6. <http://clinicaltrials.gov> (11. marts 2011).
7. Gilbody SM, Song F. Publication bias and the integrity of psychiatry research. *Psychol Med* 2000;30:252-8.
8. Altman D. The scandal of poor medical research – we need less research, better research, and research done for the right reasons. *BMJ* 1994;308:283-4.
9. Mills JL. Data torturing. *N Engl J Med* 1993;329:1196-9.
10. Sterne JAC, Smith GD, Cox DR. Stifiting the evidence – what's wrong with significance tests? *BMJ* 2001;322:226-31.

Bekæmpelse af resistente hospitalsbakterier – mindre antibiotika og mere infektionsforebyggelse

Jens Kjølsest Møller

PROFESSOR- TILTRÆDELSSES- FORELÆSNING

Klinisk Mikrobiologisk
Afdeling, Sygehus
Lillebælt, Vejle

Antibiotikaresistente hospitalsbakterier er et aktuelt, men ikke nyt problem. En sammenhæng mellem brug af antibiotika og resistensudvikling hos bakterier blev allerede varslet ved antibiotikæraens start i begyndelsen af 1940'erne. Et stærkt stigende forbrug af penicillin (eneste antibiotikum til rådighed) medførte i løbet af det følgende tiår, at hovedparten af *Staphylococcus aureus*, der var isoleret hos patienter på Londons hospitaler, producerede en penicillinnedbrydende β -laktamase [1]. Syntetisering af et β -laktamase-stabilt penicillin (meticillin) skabte en alternativ behandlingsmulighed. Penicillinresistente *S. aureus* blev imidlertid hurtigt erstattet af både penicillin- og meticillinresistente *S. aureus* (MRSA). På danske sygehuse så man en tilsvarende spredning af penicillinresistente *S. aureus* med en tiltagende udbygning af resistensmønstrene som en direkte afspejling af forbruget af antibiotika på sygehusene [2]. Brug af antibiotika selekterer resistente bakterier og skaber en *circulus viciosus* (her: resistens-ens onde cirkel), idet antibiotikaresistente bakterier forårsager behandlingssvigt, smittespredning og behandlingsforsøg med andre antibiotika (nyt selektionspres), som igen leder til selektion af hospitalsbakterier med mere komplekse resistensmønstre. Behandling af infektioner kompliceres ydermere ved, at det primære valg af det mest effektive antibiotikum ofte må erstattes af mindre effektivt virkende alternativer.

En restriktiv antibiotikapolitik blev derfor tidligt og med succes indført i Danmark med det formål, at man på samme tid effektivt kunne kurere patienter-

nes infektioner og minimere spredning af resistente mikroorganismer [2, 3].

DEN GLOBALE UDFORDRING

Antibiotikaresistens er et grænseoverskridende problem, og det er en global udfordring at udvikle nye typer af antibiotika. Skandinavien fremstår fortsat som et område med en relativt lav forekomst af resistente bakterier sammenlignet med det sydlige Europa og det meste af den øvrige verden, men antibiotikaresistente bakterier er et accelererende problem på danske sygehuse (se de årlige DANMAP-rapporter på <http://www.danmap.org>). Udviklingen afspejler et stadigt stigende antibiotikaforbrug på sygehuse og i samfundet. Globaliseringen med masseturisme inklusive sundhedsturisme, hvor patienter søger behandling i andre lande, har sammen med den udstrakte internationale handel med fødevarer skabt en global spredning af resistente bakterier fra lande med høje forekomster af antibiotikaresistens og antibiotikaforbrug til lande med lave forekomster.

De senere års beskedne udvikling af nye typer af antibiotika skaber stigende problemer med at finde alternative behandlingsmuligheder ved alvorlige infektioner med antibiotikaresistente bakterier. Aktuelt eksemplificeret ved påvisningen af en metallo- β -laktamase-producerende enterobakterie (*NDM1*), som kun er følsom over for colistin [4]. Et gammelt, nyretoksisk stof, der egentlig var betragtet som obsolet. *NDM1* er den seneste udfordring med resistente hospitalsbakterier, som i daglig omtale ofte karakteriseres med et akronym ud fra deres resistensmekanisme