

Ingezonden brief

Naar een medische seksuologie op maat. Waar is het bewijs?

Peter Leusink¹, Joan Boeke², Ellen Laan³

¹ Polikliniek Seksuologie, Groene Hart Ziekenhuis, Gouda

² Huisartsenpraktijk Postjesweg, Amsterdam

³ Afdeling Seksuologie en Psychosomatische Gynaecologie, Academisch Medisch Centrum, Universiteit van Amsterdam, Amsterdam

In 2013 werden in het *Journal of Sexual Medicine* drie artikelen gepubliceerd van dezelfde onderzoeksgroep over onderzoek naar de werkzaamheid van twee middelencombinaties bij vrouwen met verminderd seksueel verlangen (*hypoactive sexual desire disorder* of HSDD) (Poels et al., 2013; van Rooij et al., 2013; Bloemers et al., 2013). Die combinaties zijn testosteron met phosphodiesterase-5-remmer [T + PDE5i] en testosteron met buspiron [T + 5-HT_{1a}ra]. PDE5i is bekend van gebruik bij een erectiestoornis en buspiron kan worden gebruikt bij een angststoornis. De combinatie T + PDE5i wordt geacht de gevoeligheid te versterken van het brein tijdens seksuele stimulatie. De combinatie T + 5-HT_{1a}ra zou tijdelijk de toename blokkeren van remmende serotonerge activiteit in de prefrontale cortex die ontstaat als reactie op negatief ervaren seksuele stimuli (Bancroft & Janssen, 2000).

Kritische beschouwing van de drie artikelen (allem onder de titel "Naar een medische seksuologie op maat") roept veel vragen op over de validiteit van zowel de theoretische claims als van de bevindingen. De volgende onderwerpen komen aan de orde: de statistische analyse, de klinische betekenis van de bevindingen, de wijze van placebocontrole en de belangenverstengeling.

Statistische analyse

De auteurs rapporteren alleen gegevens van subgroepanalyses bij wat de auteurs noemen sterk- en zwakgeremde vrouwen (*high and low inhibitors*) en bij hoog- en laaggevoelige vrouwen (*high and low sensitive women*) (Poels et al., 2013; van Rooij et al., 2013). Deze aanpak is om meerdere redenen problematisch.

De samenstelling van de subgroep hoog- en laag-

gevoelige vrouwen is gebaseerd op scores op een zogenaamde "Emotional Stroop task". Tijdens deze taak wordt de reactietijd gemeten op erotische en neutrale woorden die in één van vier kleuren zijn afgedrukt, en die steeds worden gevolgd door een zogenaamd 'backward mask' bestaande uit *random* verwisselde letters in dezelfde kleur als het voorafgaande emotionele of het neutrale woord, er is dus geen bestaand woord leesbaar. Zo'n 'mask' verhindert dat het neutrale of erotische woord bewust kan worden waargenomen. De deelnemer moet zo snel mogelijk de kleur benoemen van de 'mask'. Relatieve ongevoeligheid voor seksuele prikkels zou samenhangen met een kortere reactietijd van het benoemen van de kleur van de 'mask' na erotische woorden dan van de kleur van de 'mask' na neutrale woorden (Poels et al., 2013, p. 815). Om de validiteit van deze indeling in subgroepen te ondersteunen, rapporteren de auteurs een toename van de automatische (onbewuste) aandacht van alleen de laaggevoelige vrouwen gedurende de actieve behandeling ten opzichte van placebo (p. 817). Te oordelen echter naar het aantal vrijheidsgraden in de ANOVA ($df = 1,54$), dat verwijst naar het totaal van 56 vrouwen in de steekproef, kan deze test niet gaan over alleen de groep laaggevoelige vrouwen (die uit 29 vrouwen bestond). Toch toont figuur 3 (Poels et al., 2013, p. 819) opnieuw deze met een foutief aantal vrijheidsgraden berekende p-waarde van 0.001. Dit roept vragen op over de nauwkeurigheid van deze analyse.

Daarnaast is het onduidelijk wat de uitkomsten waren van de Strooptest voor de subgroep van hooggevoelige vrouwen. Dat resultaat is nodig om de conclusie van de auteurs te verifiëren dat de Strooptest een goede maat is om gevoeligheid voor seksuele prikkels te meten. Mogelijk was onbewuste aandacht voor seksuele prikkels bij de T + PDE5i-behandeling niet vergroot in de hooggevoelige subgroep? Deze gegevens ontbreken in de artikelen. Overigens wordt door de auteurs de validiteit van de Strooptest ten behoeve van het vaststellen van gevoeligheid voor erotische prikkels, niet ondersteund door referenties naar ander onderzoek. Er is geen gouden standaard om laag- en hooggevoeligheid voor erotische prikkels vast te stellen.

Drs. P. Leusink, huisarts, seksuoloog NVVS

Dr. A. J. P. Boeke, huisarts

Dr. E. Laan, GZ-psycholoog, seksuoloog NVVS

Correspondentie: P. Leusink, E. leusinkp@knmg.nl

Op verzoek van het bestuur van het Tijdschrift voor Seksuologie is dit artikel een vertaling en bewerking van een ingezonden brief aan de redactie van het *Journal of Sexual Medicine* (Leusink, P., Boeke, A. J. P., & Laan, E. (2014). Toward Personalized Sexual Medicine: Where is the Evidence? *Journal of Sexual Medicine*. doi: 10.1111/jsm.12619). Beoordeling en acceptatie van dit artikel vonden plaats buiten de hoofdredacteur om.

De indeling in sterk- en zwakgeremde vrouwen is gebaseerd op dagboekscores die werden verzameld tijdens de T + PDE5i-behandeling (van Rooij et al., 2013). Vrouwen die niet goed reageerden op deze behandeling werden 'sterk-geremden' genoemd en vrouwen die wel reageerden werden 'zwak-geremden' genoemd (p. 828). Met andere woorden, remming werd gedefinieerd als gebrek aan seksuele activatie. Dit is echter in tegenspraak met het uitgangspunt van het door de auteurs gehanteerde Dual Control-model, dat de stimulerende en remmende processen, die de uiteindelijke seksuele respons mediëren, relatief onafhankelijke neurofysiologische systemen zijn (Bancroft & Janssen, 2000). Ook laten de auteurs geen andere onderzoeksbevindingen zien die dit onderscheid in subgroepen kunnen ondersteunen, zoals bijvoorbeeld andere neurofysiologische data, gevalideerde vragenlijsten of klinische bevindingen. Tabel 1 (van Rooij et al., 2013, p. 829) toont alleen enkele demografische verschillen tussen de subgroepen. Sterk-geremden hadden significant meer kinderen en relaties van langere duur dan de zwak-geremden. Juist deze contextuele variabelen hangen samen met afname van seksueel verlangen bij vrouwen (Klusmann, 2002) en met de verminderende kwaliteit van seks binnen een huwelijk (Liu, 2003). Ondanks de ingebouwde experimentele controles kunnen deze factoren niet als alternatieve verklaring voor het gebrek aan reactie op de behandeling met T + PDE5i worden uitgesloten.

Een ander issue is de onduidelijkheid over de werkelijke betekenis van de hogere prevalentie van negatieve seksuele ervaringen in de sterkgeremde groep. De auteurs veronderstellen dat negatieve seksuele ervaringen in het verleden één van de belangrijkste mechanismes is via welke vrouwen een neurofysiologisch remmingsmechanisme ontwikkelen. Vrouwen in de sterkgeremde groep hadden weliswaar meer negatieve seksuele ervaringen, maar de aard van die ervaringen varieerde sterk: van ongewenste aanrakingen tot verkrachting. We weten niet of de sterk-geremden inderdaad ook meer ernstige negatieve seksuele ervaringen hadden dan de zwak-geremden. Bovendien, ongeveer de helft van de sterk-geremden had deze ervaringen niet. Wat zou dan bij hen tot deze neurofysiologische remming hebben kunnen leiden? Tevens is het niet duidelijk in hoeverre deze ervaringen hadden geleid tot posttraumatische stress, een noodzakelijk punt van aandacht aangezien mensen verschillend op negatieve ervaringen kunnen reageren (Letourneau, Resnick, Kilpatrick, Saunders, & Best, 1996).

Nadere beschouwing van de bevindingen maakt de indeling en keuze voor de subgroepen er niet helderder op. Figuur 5 bijvoorbeeld (Poels et al., 2013, p. 820) laat zien dat T + PDE5i leidt tot een toename in gevoelens van seksueel verlangen en ervaren genitale respons ten opzichte van placebo bij de laaggevoelige vrouwen, maar niet bij de hooggevoelige vrouwen (Poels et al., 2013). Dat lijkt te passen bij de theorie,

want van hooggevoelige vrouwen verwacht je niet dat een middel dat de gevoeligheid voor seksuele prikkels verhoogt veel effect zal hebben, omdat ze al hooggevoelig zijn. De gemiddelden voor seksueel verlangen en genitale respons in Figuur 5 zijn echter ongeveer even hoog voor de laaggevoelige vrouwen op placebo als voor de hooggevoelige vrouwen op placebo. Als beide groepen al niet verschillen in seksueel verlangen en genitale respons tijdens placebo, welke disfunctie wordt er dan eigenlijk behandeld? Met andere woorden, wat is klinisch het seksuele probleem waarin beide groepen zich onderscheiden? Het lijkt alsof de auteurs de legitimiteit van de twee subgroepen baseren op de werkzaamheid van een middel, waarmee de afhankelijke en onafhankelijke variabelen feitelijk worden verwisseld.

Een goed beargumenteerde en onderbouwde verdeling in subgroepen is des te meer noodzakelijk omdat analyses van effecten van beide middelen in de gehele groep ontbreken. Waarom worden deze data niet gerapporteerd? Werden er wellicht geen significante verschillen van de middelen gevonden? Dat zou niet geheel denkbeeldig zijn, want waarom zouden de auteurs er opzettelijk voor kiezen significante bevindingen niet te rapporteren? Een recente systematische review van Sun et al. liet zien dat in trials waarbij de primaire uitkomstmaten - bedoeld om effectiviteit van een actieve behandeling ten opzichte van placebo aan te tonen - geen significante verschillen lieten zien, 2,3 keer vaker subgroepanalyses werden gerapporteerd als ze gesponsord waren door een farmaceutisch bedrijf ten opzichte van trials die niet waren gesponsord (Sun et al., 2011). Met andere woorden, wellicht moeten de subgroepanalyses in deze studie verhullen dat in de gehele groep de middelen niet verschilden van placebo.

Tevens werden de subgroepen van de hier beschreven studies niet gerandomiseerd bij aanvang van het onderzoek, een belangrijke vereiste voor betrouwbare subgroepanalyses om *confounding* van de resultaten te vermijden (Sun et al., 2010). De opvallend kleine betrouwbaarheidsintervallen binnen de relatief kleine subgroepen zouden bijvoorbeeld gewoon het gevolg kunnen zijn van de keuze om de totale groep in te delen op basis van de scores waarmee ook het effect van de behandeling zelf wordt gemeten.

Een laatste punt van kritiek ten aanzien van het statistisch-analyseplan is dat de uitkomsten niet volgens het uitgangspunt van *intention-to-treat* zijn geanalyseerd. *Intention-to-treat* wil zeggen dat na randomisatie alle uitkomsten van alle deelnemers in de data-analyse worden meegenomen, ongeacht uitvallers. Hiermee wordt voorkomen dat een gunstiger effectiviteit wordt getoond dan in werkelijkheid het geval is. Bijvoorbeeld bij geneesmiddelenonderzoek zou door het weglaten van uitvallende deelnemers met bijwerkingen de effectiviteit van de geneesmiddelen kunnen worden overschat. De door de onderzoekers gehanteerde aanpak is

niet in overeenstemming met internationaal geaccepteerde richtlijnen (Moher, Schulz, & Altman, 2001).

Klinische significantie

Conform de vereisten van de Amerikaanse Food and Drug Administration (FDA) hechtten de auteurs grote waarde aan patiëntgerichte uitkomstmaten zoals de SEI (*subjective evaluation of improvement*) en de SEG (*subjective evaluation of gain*). Beide maten zijn gemeten als de som van twee ja (1)/nee (0) vragen, resulterend in een schaal van 0 tot 2. In beide studies varieerde de toename ten opzichte van placebo binnen de subgroepen van 0.2 tot 0.4. We vragen ons sterk af wat de klinische betekenis is van een dergelijke uitkomst. Iets vergelijkbaars zien we in beide studies als vrouwen moeten rapporteren of ze seksuele verbetering ervoeren van de behandeling. In beide studies scoorden ze ongeveer 1,5 bij behandeling met het actieve middel, gemeten op een schaal van 1 tot 4 (waarbij 1 = geen, 4 = veel). Het lijkt ons niet onredelijk te veronderstellen dat dit betekent dat er nauwelijks enige verbetering zichtbaar was.

Om een goed oordeel te geven over de klinische betekenis van statistisch significante verschillen, is de effectgrootte een betere maat. Omdat in deze beide studies de effectgroottes niet zijn gerapporteerd, hebben we de effectgroottes berekend van de significante verschillen. Slechts een klein aantal effectgroottes waren groot ($0.80 < \text{Cohen's } d < 1.29$), waarvan de meeste gerelateerd aan de uitkomstenmaten SEI en SEG. Daarvan zeiden we net al dat de klinische betekenis ervan ons niet zo groot leek.

Placebocontrole

In beide empirische onderzoeken werden twee actieve behandelingen vergeleken met een placebo. Omdat er geen placebocontrole plaatsvond voor elk van de behandelingscomponenten, is het niet duidelijk welke component van de behandeling - T of PDE5i in het onderzoek dat laag- en hooggevoelige vrouwen vergeleek (Poels et al., 2013), en T of 5-HT_{1a}ra in het onderzoek dat zwak- en sterkgeremde vrouwen vergeleek (van Rooij et al., 2013) – werkelijk verantwoordelijk is voor het veronderstelde effect. Theoretisch zou de vrouw – na toelating van het medicijn – een inactieve component kunnen gaan gebruiken, hetgeen onwenselijk is met het oog op overbodig medicatiegebruik met mogelijke bijwerkingen.

Daarnaast zou het gebruik van een actieve component fysiologische bijwerkingen gehad kunnen hebben, aan de hand waarvan de vrouwen kunnen weten dat ze de actieve behandeling hebben in plaats van placebo. PDE5-remmers hebben bekende bijwerkingen als blozen, opvliegers en neusverstopping. Het is daarom noodzakelijk in onderzoeken die PDE5-remmers gebruiken ook actieve placebo's te betrekken (Feys, 2012).

Verder ontbreken in de rapportage gegevens over

hoe vaak vrouwen tijdens het onderzoek daadwerkelijk een middel hadden gebruikt. Vrouwen waren namelijk vrij om zelf te bepalen hoe vaak ze het middel wilden gebruiken. Hoeveel vrouwen gebruikten een placebo, hoeveel gebruikten het actieve middel, en beïnvloedde deze hoeveelheid hun scores op de patiëntgerichte uitkomstmaten?

Er zijn geen publicaties bekend die een positief effect laten zien op seksueel gedrag of beleving van buspiron (5-HT_{1a}ra). Er zijn ook geen gegevens over het 'zo-nodig' gebruik van dit middel. De vraag is dus gerechtvaardigd welke component in de T +5-HT_{1a}ra behandeling verantwoordelijk zou zijn voor een eventueel positief effect.

Belangenverstrengeling

In geen van de drie artikelen staat beschreven hoe rekening is gehouden met de mogelijke belangenverstrengeling van de hoofdredacteur van het Journal of Sexual Medicine, die van elk artikel medeauteur is. Wie selecteerde en beoordeelde de referenten en wie bepaalde uiteindelijk of het artikel werd geaccepteerd voor publicatie?

Daarbij hebben acht van de elf auteurs financiële banden met het farmaceutisch bedrijf dat de middelen ontwikkelt en de vier auteurs die de data analyseerden zijn eigenaar of werknemer van het farmaceutisch bedrijf dat het onderzoek financiert. Welke maatregelen om te voorkomen dat deze belangenverstrengeling invloed zou hebben op de publicatie van de artikelen?

Conclusie

Op grond van de hierboven beschreven bezwaren stellen we de vraag of de genoemde artikelen in hun huidige vorm gepubliceerd hadden mogen worden.

Literatuur

- Bancroft, J., & Janssen, E. (2000). The dual control model of male sexual response: A theoretical approach to centrally mediated erectile dysfunction. *Neuroscience and Biobehavioral Reviews*, 24, 571–9.
- Bloemers, J., van Rooij, K., Poels, P., Goldstein, I., Everaerd, W., Koppe-schaar, H., ... Tuiten, A. (2013). Toward personalized sexual medicine (Part 1): Integrating the "dual control model" into differential drug treatments for hypoactive sexual desire disorder and female sexual arousal disorder. *Journal of Sexual Medicine*, 10, 791–809.
- Feys, F., Bekkering, G. E., Singh, K., & Devroey, D. (2012). Do randomized clinical trials with inadequate blinding report enhanced placebo effects for intervention groups and nocebo effects for placebo groups? A protocol for a meta-epidemiological study of PDE-5 inhibitors. *Systematic Review*, 1, 54.
- Klusmann, D. (2003). Sexual motivation and the duration of partnership. *Archives of Sexual Behavior*, 31, 275–87.
- Letourneau, E. J., Resnick, H. S., Kilpatrick, D. G., Saunders, B. E., & Best, C. L. (1996). Comorbidity of sexual problems and post-traumatic stress disorder in female crime victims. *Behavior Therapy*, 27, 321–36.
- Liu, C. (2003). Does quality of marital sex decline with duration? *Archives of Sexual Behavior*, 32, 55–60.

- Moher, D., Schulz, K. F., Altman, D. G., & CONSORT GROUP (Consolidated Standards of Reporting Trials). (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Annals of Internal Medicine*, 134, 657–62.
- Poels, S., Bloemers, J., van Rooij, K., Goldstein, I., Gerritsen, J., van Ham, D., ... Tuiten, A. (2013). Toward personalized sexual medicine (Part 2): Testosterone combined with a PDE5 inhibitor increases sexual satisfaction in women with HSDD and FSAD, and a low sensitivity system for sexual cues. *Journal of Sexual Medicine*, 10, 810–23.
- Sun, X., Briel, M., Walter, S.,D., & Guyatt, G.,H. (2010). Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *British Medical Journal*, 340, c117.
- Sun, X., Briel, M., Busse, J.,W., You, J.,J., Akl, E.,A., Mejza, F., ..., Guyatt, G.,H. (2011). The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: Systematic review. *British Medical Journal*, 342, d1569.
- Van Rooij, K., Poels, S., Bloemers, J., Goldstein, I., Gerritsen, J., van Ham, D., ... Tuiten, A. (2013). Toward personalized sexual medicine (Part 3): Testosterone combined with a serotonin1a receptor agonist increases sexual satisfaction in women with HSDD and FSAD, and dysfunctional activation of sexual inhibitory mechanisms. *Journal of Sexual Medicine*, 10, 824–37.