

# Is standaardisatie van dierexperimenten een oorzaak van povere generaliseerbaarheid van resultaten?

F. Josef van der Staay<sup>1,3</sup>, Saskia S. Arndt<sup>2,3</sup>, Rebecca E. Nordquist<sup>1,3</sup>

FACULTEIT DIERGENEESKUNDE, PROGRAMMA EMOTIE & COGNITIE

1 DEPARTEMENT GEZONDHEIDSZORG LANDBOUWHUISDIEREN, POSTBUS 80151, 3508 TD UTRECHT  
(TELEFOON 030-2534205, E-MAIL F.J.VANDERSTAAY@UU.NL)

2 DEPARTEMENT DIER IN WETENSCHAP EN MAATSCHAPPIJ

3 RUDOLF MAGNUS INSTITUTE OF NEUROSCIENCE, UTRECHT

## Samenvatting

Het nut van standaardisatie van dierproeven is de laatste jaren herhaaldelijk in twijfel getrokken. De vraag werd gesteld of standaardisatie wellicht een oorzaak zou kunnen zijn dat sommige dierexperimentele resultaten niet goed te generaliseren zijn. De kritiek op de standaardisatie van onderzoek komt op het volgende neer: door standaardisatie worden extreem specifieke omstandigheden gecreëerd. De resultaten, verkregen met behulp van gestandaardiseerde experimenten gelden strikt genomen alleen voor deze specifieke

omstandigheden. Generalisatie naar andere condities blijkt vaak niet mogelijk. In dit artikel wordt besproken wat standaardisatie van dierexperimenteel onderzoek betekent en welke functies het heeft. We bespreken hoe we, gebruikmakend van gestandaardiseerde experimenten, verschillende wetenschappelijke vragen kunnen onderzoeken, waaronder ook de vraag of en hoe experimentele resultaten te generaliseren zijn. We stellen voor, dit doel te bereiken m.b.v. een 'replicatie'-strategie. Deze strategie wordt uitvoerig besproken.

## Inleiding

Bijna ieder van ons zal in zijn/haar opleiding geleerd hebben dat herhaalbaarheid van resultaten van experimenten een van de hoekstenen van wetenschappelijk onderzoek is. Daarom bevatten alle relevante proefdierkundige boeken hoofdstukken over het plannen en uitvoeren van dierexperimenten volgens strikte protocollen (bijvoorbeeld 1, 2). Door standaardisatie kunnen de resultaten van studies afkomstig uit verschillende laboratoria met elkaar worden vergeleken. Bovendien kunnen de resultaten worden vergeleken met uitkomsten van eerder eigen onderzoek (historische data). Dit is vooral van belang bij het onderzoek naar teratologische en toxicologische risico's bij de ontwikkeling van nieuwe geneesmiddelen.

Het nut van standaardisatie is de laatste jaren herhaaldelijk in twijfel getrokken (3-6). De vraag werd gesteld of standaardisatie wellicht een oorzaak zou kunnen zijn van de povere generaliseerbaarheid van dierexperimentele resultaten (3). De kritiek op standaardisatie van onderzoek komt op het volgende neer: door standaardisatie worden extreem specifieke omstandigheden gecreëerd. De verkregen resultaten gelden strikt genomen alleen voor deze specifieke omstandigheden (idiosyncratische resultaten). Omdat ze niet generaliseerbaar zijn voor andere condities kunnen dergelijke resultaten niet eenduidig als gevolgen van de experimentele manipulatie worden beschouwd.

Recent werd een oplossing voor dit probleem voorgesteld (3, 4). In plaats van standaardisatie zou juist een heterogene omgeving moeten worden toegepast ('heterogenisatie'). Immers, als we dieren in verschillende omstandigheden houden en een bepaalde experimentele manipulatie dezelfde effecten geeft dan zijn de resultaten niet idiosyncratisch. Anders gezegd: de effecten doen zich voor onder verschillende omstandigheden en zijn daardoor generaliseerbaar. Deze kritiek op de gebruikelijke standaardisatie van dierexperimenteel onderzoek heeft bij velen geleid tot onzekerheid en vragen.

*Over standaards en standaardisatie bestaan veel misvattingen. We bespreken in de volgende paragrafen wat standaardisatie is, of het klopt dat resultaten uit gestandaardiseerde studies een beperkte generaliseerbaarheid hebben, of 'heterogenisatie' een alternatieve aanpak van het generalisatieprobleem is, en of er andere alternatieven zijn om dit probleem op te lossen.*

### **Standaards en standaardisatie**

Wat is standaardisatie, en wat is het niet? Het is niet makkelijk, een goede definitie van standaardisatie te geven, maar de volgende is erg bruikbaar:

*Standaardisatie is het ontwikkelen en vastleggen van technische standaards in een document dat uniforme specificaties, criteria, methoden, processen en praktijken beschrijft. Standaardisatie is in wezen op consensus gebaseerd, en streeft ernaar een hoog niveau van herhaalbaarheid, vergelijkbaarheid, compatibiliteit en kwaliteit van een experiment te waarborgen, en een valide vergelijking tussen onderzoeken mogelijk te maken.*

Voor dierexperimenteel onderzoek betekent dit, dat men vooraf kenmerken van experimentele factoren vastlegt, zoals de opgroei-, huisvestings- en testomstandigheden van het proefdier, en dat men de testapparatuur en de testmethoden nauwkeurig beschrijft.

### **Verkeerd gebruik van het begrip standaardhuisvesting**

Sommige auteurs, en dit geldt ook voor Richter en collegae, die 'heterogenisatie' voorstellen, associëren het begrip 'standaardhuisvesting' met groepshuisvesting

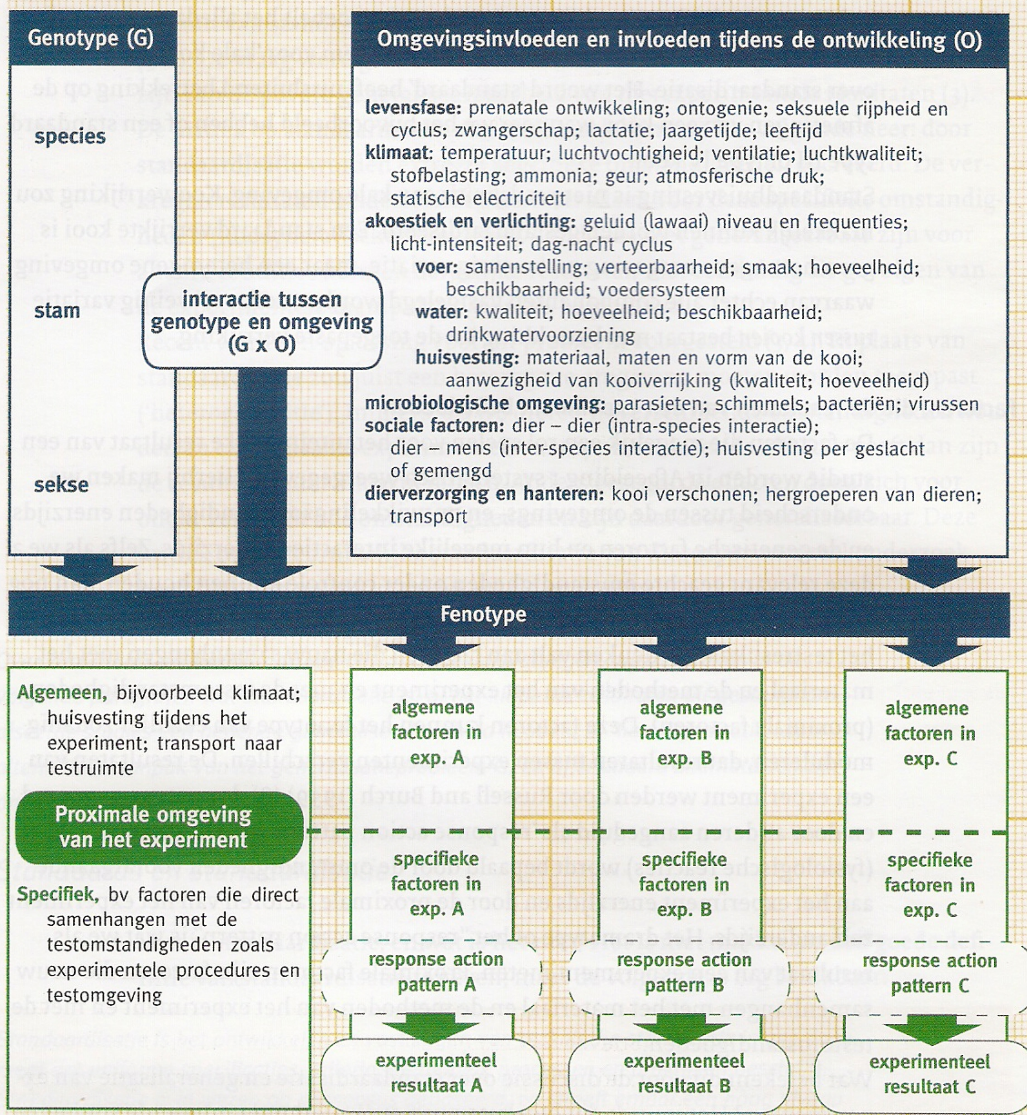
in kooien zonder kooiverrijking (7). Helaas vertroebelt het alledaagse gebruik van het begrip 'standaardhuisvesting' als synoniem voor 'kale kooi' de discussie over standaardisatie. Het woord 'standaard' heeft uitsluitend betrekking op de afmetingen van een kooi, wanneer we het bijvoorbeeld hebben over een standaard type III of type IV kooi.

Standaardhuisvesting is niet *per definitie* een kale omgeving. Kooiverrijking zou makkelijk kunnen worden gestandaardiseerd. Een standaard verrijkte kooi is geen homogene omgeving met weinig variatie, maar een heterogene omgeving waarvan echter alle componenten vastgelegd worden, zodat er weinig variatie tussen kooien bestaat met betrekking tot de toegepaste verrijking.

### **Factoren, die de resultaten van een experiment beïnvloeden**

De factoren die mogelijk een rol spelen voor het uiteindelijke resultaat van een studie worden in Afbeelding 1 systematisch weergegeven. Hierbij maken we onderscheid tussen de omgevings- en ontwikkelingsomstandigheden enerzijds, en de genetische factoren en hun mogelijke interacties anderzijds. Zelfs als we al deze relevant geachte omstandigheden onder controle kunnen houden, dan nog kunnen ongecontroleerde factoren tijdens het onderzoek invloed uitoefenen op het testresultaat. Te denken valt aan factoren die nauw samenhangen met het materiaal en de methoden van het experiment en met de testomstandigheden (proximale factoren). Deze factoren kunnen het fenotype van een dier zodanig moduleren, dat resultaten tussen experimenten verschillen. De resultaten van een experiment werden door Russell and Burch (1959) (8) dramatype genoemd en door anderen aangeduid als 'response action pattern' (9). Dit patroon van (fysiologische reacties) wordt bepaald door de omstandigheden voorafgaande aan het experiment enerzijds en door de proximale factoren van het experiment zelf anderzijds. Het dramatype of het "response action pattern" is wat we als resultaat van een experiment meten. Proximale factoren zijn factoren die nauw samenhangen met het materiaal en de methoden van het experiment en met de testomstandigheden.

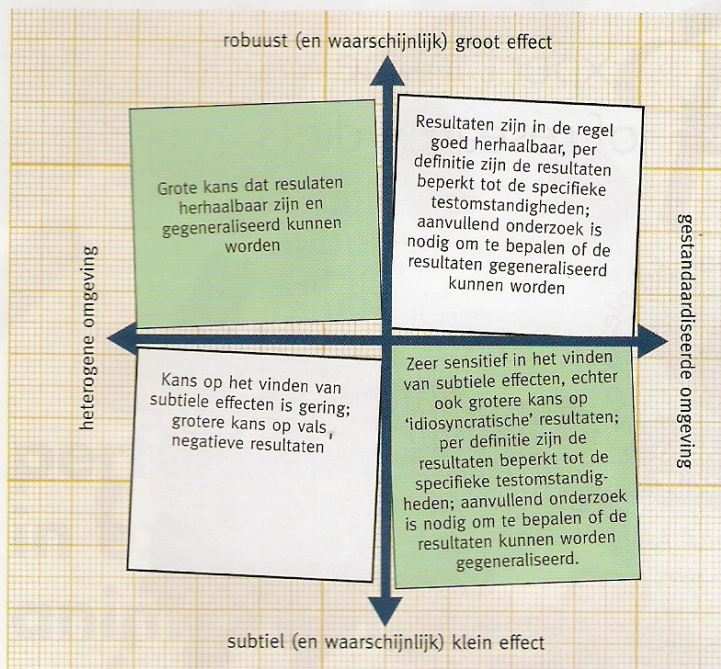
Wat betekent dit voor de discussie over standaardisatie en generalisatie van experimenten? Het is waar dat resultaten van strikt gestandaardiseerde studies in principe alleen gelden voor de condities, waaronder de studie werd uitgevoerd. In Afbeelding 1 zien we dat een groot aantal factoren een rol kan spelen bij wat we uiteindelijk in een experiment meten. Deze factoren kunnen subtiele interacties met elkaar aangaan, zodanig dat we geen duidelijk effect van een experimentele manipulatie kunnen detecteren. Dit kan bijvoorbeeld gebeuren als niet nader geïdentificeerde omgevingsinvloeden, op individuele dieren tegengestelde effecten hebben. De waarschijnlijkheid is dus groot dat we bij heterogenisatie vooral robuuste (en sterke effecten) opmerken en subtielere effecten over het hoofd zien (Afb. 2).



Afbeelding 1.

Het genotype, omgevings- en ontwikkelingsinvloeden en hun interacties (blauw) bepalen het fenotype van de modeldieren, die we in experimenten inzetten (11, naar 12). In experimenten die volgens het principe heterogeniseren worden uitgevoerd, wordt een breed spectrum van omgevingsfactoren – al dan niet systematisch – gevarieerd. Sommige van de omgevingsfactoren hebben waarschijnlijk geen invloed op het resultaat van het experiment, maar andere kunnen het resultaat beïnvloeden. Helaas is niet altijd bekend, welke factoren dat zijn.

Daarom zijn systematische, zorgvuldig gecontroleerde experimenten nodig om te onderzoeken, welke (omgevings)factoren de experimentele resultaten beïnvloeden. In het onderste gedeelte van deze afbeelding (groen) is het concept van 'response action pattern' systematisch weergegeven. Algemene en specifieke proximale factoren kunnen de resultaten van een experiment beïnvloeden. Het moge duidelijk zijn dat de proximale factoren tot verschillende experimentele resultaten kunnen leiden, zelfs wanneer we met een strikt gedefinieerd en gecontroleerd fenotype werken.



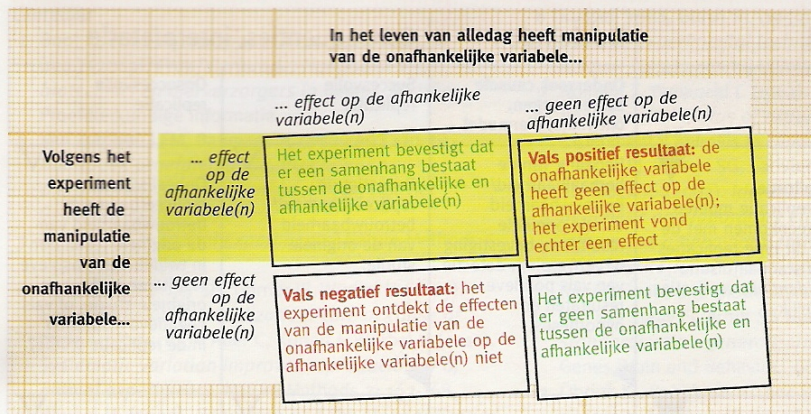
Afbeelding 2.  
 We vermoeden dat de twee dimensies heterogeniseren vs. standaardiseren en robuuste vs. subtiel effecten niet onafhankelijk van elkaar zijn. Onder de condities van heterogenisatie vinden we vooral robuuste effecten. De kans op het opsporen van subtiel effecten is het grootst in een gestandaardiseerde omgeving.

### Waarom kunnen subtiel experimentele effecten belangrijk zijn?

Het opsporen van subtiel effecten kan van groot belang zijn. Een team wil bijvoorbeeld met behulp van een geschikt diermodel onderzoeken of een nieuwe therapeutische aanpak mogelijk effectief is. Om de mogelijke klinische effectiviteit te toetsen staan nog geen substanties met hoge selectiviteit ter beschikking, maar er zijn wel substanties die een zekere activiteit hebben aan de receptor die betrokken zou kunnen zijn bij het beoogde therapeutische effect. Wanneer deze testsubstanties alleen onder zeer specifieke omgevingsomstandigheden effect hebben, dan lopen we in gestandaardiseerde en in geheterogeniseerde experimenten het risico op een fout-negatieve uitkomst. Het concept van fout-positieve en fout-negatieve uitkomsten van een experiment is in Afbeelding 3 weergegeven. Bij gestandaardiseerde experimenten kan men als volgende stap systematisch onderzoek uitvoeren naar de vraag, onder welke omstandigheden een klinische werking optreedt. Op deze optie komen we in de volgende paragraaf uitvoeriger terug. Het gevolg van negatieve uitkomsten van de heterogenisatie-aanpak zou zijn dat verder onderzoek wordt gestaakt. Het identificeren van de 'exacte gen bij omgeving' interacties (Afb. 1) kan van het grootste belang zijn voor de ontwikkeling van geïndividualiseerde farmacotherapieën (10).

### Meertrapse replicatiebenadering

Welke alternatieve benadering is er om het probleem standaardisatie – generalisatie op te lossen? We hebben gezien dat 'heterogenisatie' kan worden toegepast om een indruk te krijgen van de generaliseerbaarheid van experimentele resul-



Afbeelding 3. De waarschijnlijkheid van vals negatieve resultaten hangt onder andere af van de gevoeligheid van de experimentele methoden en de grootte van het effect.

taten. We hebben ook gezien dat men met deze benadering kans loopt, subtiele, maar belangrijke effecten te missen. Een meertrapse replicatiebenadering is in staat, onder behoud van de voordelen van gestandaardiseerde experimenten (bepaling van repliceerbaarheid/ betrouwbaarheid), de vraag naar de generaliseerbaarheid van experimentele effecten te beantwoorden. Deze benadering kan gevoelig subtiele effecten detecteren en is in Afbeelding 4 schematisch weergegeven. De verschillende typen replicatiestudies, de doeleinden waarvoor ze worden toegepast en de interpretatie van succesvolle en onsuccesvolle replicaties zijn in de afbeelding uitgewerkt.

Na de originele studie wordt er een eerste exacte of directe replicatie uitgevoerd, die vooral tot doel heeft de herhaalbaarheid en betrouwbaarheid van het originele onderzoek te bepalen. Er kan een reeks uitgebreidere 'partiële', 'systematische', 'differentiële' of 'conceptuele' replicaties volgen, die ieder een ander aspect van het experiment variëren. Met behulp van deze replicaties wordt de generaliseerbaarheid van resultaten aan een systematische toetsing onderworpen.

Deze reeks van replicaties kan worden afgesloten met een 'quasireplicatie' waarin een ander modeldier dan in de eerdere studies wordt ingezet. Preklinisch onderzoek naar de efficiëntie, maar vooral de veiligheid van potentiële geneesmiddelen, wordt vaak met op zijn minst twee verschillende species uitgevoerd. Wanneer deze tests vergelijkbare resultaten opleveren, d.w.z. dat resultaten over species heen kunnen worden ggeneraliseerd, kan dit de generalisatie naar de mens vergemakkelijken. De kans op succesvolle generalisatie is groter naarmate de tweede species fylogenetisch dichter bij de mens staat, d.w.z. naarmate ze vanuit de evolutie gezien nauwer verwant is met de mens. Ook kan het onderzoek met een andere species dan in het oorspronkelijke onderzoek een nieuw diermiddel opleveren (14).

Replicatie batterij	Doeleinden	Consequenties, interpretatie	
<b>Originele studie</b>	<b>Onderzoek causale samenhangen; ontwikkel diemodel</b>	Succesvolle replicatie	Onsuccesvolle replicatie
<i>Eerste replicatie(s)</i> <ul style="list-style-type: none"> <li>• <b>'Exacte' of 'directe' replicatie:</b> alle aspecten van de replicatie moeten zo sterk mogelijk overeenkomen met de originele studie. Voor dit type replicaties is een grote mate van standaardisatie noodzakelijk. Echter, <i>identieke</i> replicaties zijn nagenoeg onmogelijk.</li> </ul>	Bepaling van de herhaalbaarheid/betrouwbaarheid van de originele resultaten; bevestiging of verwerping; detectie van vals positieve resultaten en hypothesen.	Bevestigt de repliceerbaarheid/betrouwbaarheid van de originele studie.	Onbeslist; trekt repliceerbaarheid/betrouwbaarheid van de originele resultaten in twijfel; vals positieve resultaten in de originele studie, of vals negatieve resultaten in de replicatie studie?
<i>Volgende, uitgebreide replicaties</i> <ul style="list-style-type: none"> <li>• <b>'Partiële replicatie:</b> enkele procedurele modificaties, terwijl alle andere aspecten gelijk zijn aan die van de originele studie.</li> <li>• <b>'Systematische' of 'differentiële' replicatie:</b> variaties in de belangrijkste onafhankelijke variabelen (bv opgroei- huisvestings- en testcondities, geslacht) welke uitgebreid kunnen worden tot semi-natuurlijke omgevingen;</li> <li>• <b>'Conceptuele' replicatie:</b> onderzoek naar dezelfde samenhangen, constructen als in de originele studie, maar met behulp van andere procedures.</li> </ul>	Bepaling van de generaliseerbaarheid (externe validiteit); identificeer de omstandigheden waaronder generaliseerbaarheid niet mogelijk is; identificeer mogelijk versturende variabelen.	Bevestigt de generaliseerbaarheid van resultaten; levert soms nieuwe inzichten op die een nieuwe ronde van het formuleren en testen van aangepaste of nieuwe hypothesen op gang zetten.	Ontdekt condities waaronder de generaliseerbaarheid niet geldt; toont de grenzen van de generaliseerbaarheid van resultaten van de originele studie.
<i>Quasireplicatie(s)</i> <ul style="list-style-type: none"> <li>• <b>Andere species dan in de originele studie wordt getest</b></li> </ul>	Bepaling of resultaten gegeneraliseerd (externe validatie) kunnen worden naar andere species; bepaling van de vertaalbaarheid van resultaten naar andere species.	Bevestigt de generaliseerbaarheid van resultaten naar andere species, steunt de 'translationele waarde' van de resultaten.	Trekt in twijfel dat de resultaten kunnen worden gegeneraliseerd naar andere species.

Afbeelding 4.

Een meertraps benadering van replicatiestudies kan worden toegepast om de herhaalbaarheid, betrouwbaarheid (interne validiteit) en de mate waarin experimentele resultaten kunnen worden gegeneraliseerd (externe validiteit) systematisch te bepalen (naar 9, 11, 13).

## Conclusies

Heterogenisatie is een methode om de generaliseerbaarheid van resultaten te onderzoeken (3). Deze methodologische benadering heeft echter een aantal nadelen die we boven hebben beschreven. Wij geven de voorkeur aan een systematische benadering, die gebaseerd is op een reeks replicatiestudies. Met behulp van deze benadering kan een groter aantal wetenschappelijke vragen worden beantwoord dan bij experimenten die gebaseerd zijn op heterogenisatie. De systematische, replicatie-gebaseerde benadering kan subtiele effecten detecteren, helpt bij het identificeren van genetische en omgevingseffecten en hun interacties, en levert data voor de extrapolatie en generalisatie van resultaten op.

## Literatuur

- 1 Boot R, Prins J-B, Rooymans TP, & Strootman, JC (red.) (2009). *Proefdierkunde. Handboek voor biotechnici en diervverzorgers* (2 delen), Stichting Proefdierkundige Informatie.
- 2 Van Zutphen LFM, Baumans V, Ohl F (ed.) (2009). *Handboek proefdierkunde. Proefdieren, dierproeven, alternatieven en ethiek*, Elsevier Gezondheidszorg, Maarssen.
- 3 Richter lity in animal experiments? *Nature Methods* 6: 257-261.
- 4 Richter SH, Garner SH, Garner JP, Würbel H (2009). *Environmental standardization: cure or cause of poor reproducibility*, Auer C, Kunert J, Würbel H (2010). *Systematic variation improves reproducibility of animal experiments*. *Nature Methods* 7: 167-168.
- 5 Würbel H. (2000). *Behaviour and the standardization fallacy*. *Nature Genetics* 26: 263.
- 6 Würbel H (2002). *Behavioral phenotyping enhanced – beyond (environmental) standardization*. *Genes Brain and Behavior* 1: 3-8.
- 7 Götz N, Lewejohann L, Tomm M, Ambrée O, Keyvani K, Paulus W, Sachser N (2008). *Effects of environmental enrichment on exploration, anxiety, and memory in female TgCRND8 Alzheimer mice*. *Behavioural Brain Research* 891: 43-48.
- 8 Russell WMS, Burch RL (1959). *The principles of humane experimental technique*, Methuen; in herdruk verschenen bij UFAW, 1992.
- 9 Nomura T, Yamauchi C (1964). *Environments and physiological status of experimental animals*. *Experimental animals in cancer research*, pp. 17-35.
- 10 Chesler EJ, Wilson SG, Lariviere WR, Rodriguez-Zas SL, Mogil JS (2002). *Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive*. *Neuroscience and Biobehavioral Reviews* 26: 907-923.
- 11 van der Staay FJ, Arndt SS Nordquist RE. *The standardization-generalization dilemma: a way out*. *Genes Brain and Behavior*. (in druk).
- 12 Öbrink KJ, Rehlinger C (2000). *Animal definition: a necessity for the validity of animal experiments?* *Laboratory Animals* 34: 121-130.
- 13 Rosenthal R (1991). *“Replication in behavioral research”* in *Replication research in the social sciences*, ed. J.W. Neuliep, Sage Publications, Newbury Park, pp. 1-30.
- 14 Van der Staay FJ, Arndt SS, Nordquist RE (2009). *Evaluation of animal models of neurobehavioral disorders*. *Behavioral and Brain Functions*, pp 5, 11, 23

All it takes to operate

# B | BRAUN

## AESCULAP

- Aesculap instrumentarium
- Hechtmateriaal en wondverzorging
- Desinfectie, hygiëne en overige bescherming
- Spuiten, naalden, I.V. canules en infuuspompen
- Producten voor bereiding en toediening infusie vloeistoffen
- Water voor injectie en andere infusie vloeistoffen in diverse presentaties



Bio Services  
www.bio-services.nl

Bio Services is uw dealer voor een groot assortiment B. Braun producten



hechtdraad voor alle diersoorten



hechtapparaat voor alle diersoorten

Eén adres voor kwaliteit, continuïteit en service

Postbus 29 5400 AA Uden  
T +31 (0) 413 20 50 30 F +31 (0) 413 20 50 39  
E info@bio-services.nl



Bio Services  
www.bio-services.nl