


Vermindering van het aantal proefdieren in dierstudies



Eric Schoen*,
Suzanne Spaan,
Pascalie van Loo en
Cyrille Krul

TNO, Postbus 360, 3700 AJ Zeist. *e-mail: eric.schoen@tno.nl; telefoon: 0 888 666 347.

In dit artikel worden concrete opties besproken om het proefdiergebruik in regulatoire studies en niet regulatoire studies in de pre-klinische fase terug te brengen via statistische methoden. Voor regulatoire studies, bijvoorbeeld naar de effectiviteit van batches geneesmiddelen, bepleiten we de invoering van databases met historische controlegroepen. Hierdoor hoeft er niet in elke studie opnieuw een controle groep te worden meegenomen. Voor niet regulatoire studies, bijvoorbeeld naar de werkzaamheid van potentiële geneesmiddelen, laten we zien dat het voordelen biedt om verhoudingsgewijs meer dieren aan de negatieve controlegroep dan aan de behandelingsgroepen toe te kennen.

Oude leermeesters

Dit artikel gaat over de mogelijkheden die statistische methoden bieden om het aantal proefdieren in dierstudies terug te dringen. Die methoden spelen een cruciale rol in de evaluatie van de studies en kunnen naar ons idee vaak leiden tot proefdierversmindering. Graag citeren we in dit verband uit het boek van Russell en Burch (6): 'Every time any particle of statistical method is properly used, fewer animals are employed than otherwise would have been necessary'. Dit citaat stamt uit 1959, maar de onderliggende aansporing daagt nog steeds uit.

Vergelijkingen tussen groepen

Dierstudies bestaan meestal uit meerdere experimentele groepen. Aan het begin worden de proefdieren willekeurig aan die groepen toegekend. Na afloop van de studie gaat men met statistische toetsen na of biologisch relevante parameters in de ene groep verschillen van die van een andere groep. Ook kunnen meerdere groepen met meerdere andere groepen worden vergeleken. De gemaakte vergelijkingen hangen van het doel van de studie af, maar in de meeste gevallen gaat het om vergelijkingen tussen een groep met onbehandelde dieren, de controlegroep, en een groep dieren die een bepaalde behandeling krijgen.

De statistische toetsen waarmee groepen worden vergeleken zijn hulpmiddelen om te bepalen »

Statistische toets	Werkelijkheid	
	Geen verschil	Wel verschil
Geen verschil	Correct	Fout negatief
Wel verschil	Fout positief	Correct

Tabel 1. Correcte en foutieve conclusies uit een statistische toets.

of die groepen meer dan door toevallige oorzaken van elkaar verschillen. In Tabel 1 is een overzicht gegeven van de mogelijke conclusies uit een statistische toets. In de eerste plaats is het mogelijk dat er in werkelijkheid geen verschil is en dat de statistische toets ook geen verschil laat zien. In dat geval trekken we dus een correcte conclusie. Het is ook mogelijk dat de twee groepen puur door toeval van elkaar verschillen. Als we vervolgens concluderen dat het verschil door onze interventie tot stand is gekomen, dan trekken we een fout positieve conclusie. Een essentieel onderdeel van elke statistische toets is de mogelijkheid om de kans op een fout positieve conclusie te beheersen. We kunnen bijvoorbeeld specificeren dat we niet meer dan 5% kans op zo'n conclusie willen lopen.

Power

Als er in werkelijkheid wel een verschil tussen groepen bestaat, kan dit door toevalsvariatie zo gemaskeerd worden dat dit niet is terug te zien in het resultaat van de statistische toets. Als we dan concluderen dat er geen verschil is, dan is dat een fout negatieve conclusie. We willen graag dat de kans op zo'n conclusie tot een aanvaardbaar niveau wordt teruggedrongen. Anders gezegd: we willen dat de *power* van statistische toetsen boven een minimaal aanvaardbare waarde komt te liggen.

We kunnen de power laten toenemen door het aantal dieren in de studie te verhogen. Deze optie verdraagt zich slecht met het gedachtegoed van Russell en Burch (6). Vanuit de huidige wetgeving met betrekking tot dierproeven en vanuit financieel oogpunt is dit evenmin een goede optie. In dit artikel gaan we na of het mogelijk is om de power van een studie te laten toenemen terwijl het aantal gebruikte dieren gelijk blijft of zelfs afneemt. Om de mogelijkheden in kaart te brengen, maakten we een tweedeling van het dierexperimenteel onderzoek volgens de vrijheid die de onderzoekers in de proefopzet hebben bij regulatoire en niet regulatoire studies. Van elk type dierstudie bespreken we vervolgens de mogelijkheden om tot een vermindering van proefdiergebruik te komen.

Vrijheid bij de opzet van dierstudies

Regulatoire studies moeten worden uitgevoerd volgens richtlijnen die zijn vastgelegd in wet- en regelgeving. Deze studies zijn er meestal op gericht de toxiciteit van een teststof vast te stellen. De richtlijnen specificeren minimale aantallen dieren per groep. Dieren in een experimentele groep worden aan één van een aantal mogelijke doseringen van de teststof blootgesteld. Er is een negatieve controle om het normale niveau van een studieparameter voor het toxische effect te bepalen. In sommige gevallen is er ook een positieve controlegroep om vast te stellen of een toxisch effect wel gemeten kan worden. De groepen zijn allemaal even groot.

Voor de niet-regulatoire studies beperken we ons tot studies die erop gericht zijn de werkzaamheid van een potentieel geneesmiddel aan te tonen in diermodellen waarin een ziekte is geïnduceerd. Onderzoekers bepalen in dit soort studies zelf de opzet. Wel kan er sprake zijn van randvoorwaarden die door een regelgevende instantie, bij voorbeeld de Amerikaanse Food and Drugs Administration (FDA), worden opgelegd. Het middel wordt in diverse doseringen onderzocht. Ook is er een negatieve controlegroep (wel ziekte, geen of placebo behandeling) om het effect van de doseringen mee te vergelijken.

In de volgende paragrafen bespreken we voor beide groepen studies de mogelijkheden om tot vermindering van het aantal proefdieren te komen.

Regulatorische studies – historische controles

Op het eerste gezicht lijkt er weinig ruimte te zijn om het aantal proefdieren in een regulatorische studie te verminderen. Er is immers voorgeschreven hoeveel dieren er in de doseringsgroepen gebruikt dienen te worden. De richtlijnen schrijven echter niet voor welke controlegroepen exact moeten worden meegenomen. Zo suggereerden deelnemers van een workshop om proefdiergebruik in genotoxiciteitsstudies te verminderen door de positieve controlegroepen in de micronucleustesten niet in elke test mee te nemen (4). In plaats daarvan zou men van tijd tot tijd, bijvoorbeeld elk half jaar, een positieve controlegroep kunnen opnemen in een monitorprogramma en deze toevoegen aan de data uit de historische controles.

Een alternatief voor het volledig loskoppelen van de positieve controlegroep is de instelling van kleine controlegroepen voor elke studie met een gelijktijdige opbouw van een historische database. Elke nieuwe controlegroep zou dan vergeleken kunnen worden met de groepen die al in de database zitten. Als de gegevens van de nieuwe groep in lijn zijn met wat men op grond van de historische data kan verwachten, dan worden die gegevens aan de database toegevoegd en kan men de andere proefgroepen van de studie afzetten tegen de gezamenlijke positieve controlegroepen. Indien de gegevens niet in lijn zijn met de historische controle, dan is dit, net als bij gebruik van een volledige positieve controlegroep, een aanwijzing dat de studie niet valide is.



Verschillen tussen de sexen

In de workshop die we in de vorige paragraaf noemden (4) werd ook aanbevolen om studies niet per definitie zowel met mannetjes als met vrouwtjes uit te voeren, maar alleen met de meest gevoelige sexe. Als het onduidelijk is welke sexe het meest gevoelig is voor de stof die men wil onderzoeken, dan kunnen gelijke aantallen dieren per sexe worden ingezet, maar dan met in totaal even veel dieren als wanneer men één sexe zou onderzoeken. »

Providing Quality Purpose Bred Animals for Biomedical Research



 *Beagles now available in the UK*




BIORESOURCES



- Beagle Production in the United States, Europe and China
- Comprehensive Socialization and Enrichment Programs
- Dedicated to Flexible, Personalized Customer Service
- ISO-9001 Certified Quality Systems

 North America
+1 315.587.2295
infous@marshallbio.com

 Europe
+33 4 72 56 98 60
infoeu@marshallbio.com

 Asia
+86 10 84923662
infoch@marshallbio.com

Tabel 2. Proefopzet voor een regulatoire studie met twee sexen en drie doseringen. De dubbele pijlen geven aan welke vergelijkingen worden gemaakt.

		sexe		dosering				
		controle	laag	midden	hoog			
OUD	♂	8	8	8	8	Twee aparte analyses		
		←————→						
		←————→						
		←————→						
	♀	8	8	8	8	Eén gezamenlijke analyse		
	←————→							
	←————→							
	←————→							
NIEUW	♂	4(Cm)	4(Lm)	4(Mm)	4(Hm)	Eén gezamenlijke analyse		
		←————→						
		←————→						
		←————→						
	♀	4(Cv)	4(Lv)	4(Mv)	4(Hv)	Eén gezamenlijke analyse		
	←————→							
	←————→							
	←————→							

*In het groene gedeelte staan tussen haakjes de groepsaanduidingen vermeld: C = controle, L = lage dosering, M = midden dosering, H = hoge dosering, m = mannetjes, v = vrouwtjes.

Deze optie voor vermindering, die uiteraard alleen mogelijk is indien de te volgen richtlijnen hier de ruimte voor bieden, is weergegeven in Tabel 2. We gaan er in dit voorbeeld van uit dat de richtlijnen acht dieren per groep specificeren.

Normaliter zouden aparte statistische analyses worden uitgevoerd voor de mannetjes en voor de vrouwtjes. Door gebruik te maken van een *variantieanalyse voor twee studiefactoren* (7; hoofdstuk 11) kan men de verschillen tussen dosering en controlegroepen in de gecombineerde opzet met eenzelfde nauwkeurigheid testen als met de afzonderlijke toetsen voor mannetjes en vrouwtjes in de oude opzet. Voor de vergelijking van de middendosering en de controle vergelijkt men beide middendoseringsgroepen (Mm en Mv) met beide controlegroepen (Cm en Cv), dus acht dieren versus acht dieren. In formulevorm kijken we naar $(Mm+Mv) - (Cm+Cv)$. In deze formule staat Mm voor het gemiddelde van de middendoseringsgroep voor mannetjes, Mv voor het gemiddelde van de middendoseringsgroep voor vrouwtjes, en Cm en Cv voor de gemiddelden van de controlegroepen voor respectievelijk mannetjes en vrouwtjes.

Om na te gaan of het verschil tussen middendosering en controle bij vrouwtjes anders is dan bij mannetjes, berekenen we $(Mv - Cv) - (Mm - Cm)$, of wel $(Mv + Cm) - (Cv + Mm)$. Opnieuw vergelijkt men dan acht dieren met acht andere dieren. Door deze methodiek wordt het aantal proefdieren met 50% verminderd.

Recovery controlegroepen

Konigsson, Robinson en Harlemann bespreken éénmaands toxicologische studies met honden (3, 5). Een algemeen toegepaste opzet voor deze studies is weergegeven in Tabel 3. Er is zowel een recovery controlegroep als een recoverygroep voor de hoogste dosering. De laatstgenoemde recoverygroep geeft antwoord op de vraag of een effect blijvend is of weer hersteld kan worden. De auteurs zetten vraagtekens bij de recovery controlegroep omdat deze groep geen antwoord geeft op herstel na optreden van een effect. Door de recovery controlegroepen achterwege te laten verminderen ze het aantal proefdieren met 6-30 per studie. »

Tabel 3. Proefopzet voor twee sexen en drie doseringen met recoverygroepen. De dubbele pijlen geven aan welke vergelijkingen worden gemaakt.

	sexe	dosering			
		controle	laag	midden	hoog
HOOFDDEEL	♂	8	8	8	8
	♀	8	8	8	8
		↕??			↕
RECOVERYDEEL	♂	3			3
	♀	3			3

Regelgevende instanties hebben deze interpretatie van de richtlijnen geaccepteerd voor toxiciteitsstudies met honden. Indien deze aanpak ook in studies met knaagdieren kan worden doorgevoerd, levert dit een verveelvoudiging van de besparing op. Tachtig procent van alle toxiciteitsstudies wordt namelijk met knaagdieren uitgevoerd – al hebben lang niet al deze studies twee recoverygroepen.

Niet-regulatorische studies naar werkzaamheid van een geneesmiddel

Dierstudies naar werkzaamheid van een potentieel nieuw geneesmiddel omvatten in het algemeen een negatieve controlegroep, meerdere doseringsgroepen en een positieve controlegroep. In de negatieve controlegroep induceert men een bepaalde aandoening bij de dieren die niet met een geneesmiddel of met een placebo wordt behandeld. In de doseringsgroepen hebben de dieren die aandoening ook gekregen, maar worden ze behandeld met de stof die onderzocht wordt. De dieren in de positieve controlegroep hebben de aandoening eveneens gekregen, maar ze zijn behandeld met een geneesmiddel waarvan bewezen is dat het de aandoening wegneemt.

De positieve controle moet aantonen dat de proefdieren de potentie hebben om te herstellen van de aandoening. Als dat zo is, dan beschouwt men de studie als valide. Het doel van de studie is echter de vergelijking tussen de groepen behandeld met het nieuwe geneesmiddel en de negatieve controle (wel ziekte, geen geneesmiddel).

Voorbeeld met drie doseringsgroepen

Als voorbeeld gaan we ervan uit dat er naast de negatieve controlegroep groepen zijn met drie doseringen van een geneesmiddel. We bekijken twee opties voor het ontwerp van het experiment:

1. Aan elke proefgroep kennen we een gelijk aantal dieren toe.
2. We zorgen ervoor dat het aantal dieren in de controle groep $\sqrt{3}$ maal groter is dan het aantal dieren in een behandelgroep.

In de eerste optie omvat elke proefgroep 25% van het totaal aantal dieren. In de tweede optie omvat de controlegroep 37% van het totaal aantal dieren en omvat elk van de behandelgroepen 21% van het totaal. Vergelijkingen tussen de behandelgroepen onderling worden hiermee onnauwkeuriger, maar vergelijkingen van de behandelgroepen met de controlegroep worden juist nauwkeuriger. De aangegeven percentages volgen uit een berekening voor de optimalisatie van de nauwkeurigheid van de vergelijkingen. Hieruit blijkt dat bij k behandelgroepen het aantal dieren in de controlegroep \sqrt{k} groter moet zijn dan in een behandelgroep (1, 2).

Minder dieren bij ongelijke groepsgrootte

In afb. 1 laten we het aantal dieren zien dat voor de twee bovenstaande opties nodig is om een effect van een bepaalde grootte te kunnen aantonen. De bovenste lijn, in rood, heeft betrekking op een gelijke verdeling van dieren over de groepen (optie 1) en de onderste lijn, in groen, op een ongelijke verdeling van de groepen (optie 2).

De effectgrootte op de X-as is uitgedrukt in standaarddeviaties. Die standaarddeviaties drukken op hun beurt uit hoe twee dieren in eenzelfde groep kunnen verschillen wat betreft een eindpunt van de studie. Laten we als voorbeeld aannemen dat onze interesse uitgaat naar huiddikte, en dat de standaarddeviatie van huiddikte tussen dieren van een groep 0,9 mm bedraagt. In dat geval kunnen twee dieren alleen al op grond van toeval $2\sqrt{2}$ maal de standaarddeviatie, dus 2,55 mm, in huiddikte verschillen.

Stel dat men een studie wil starten om een vermindering van de huiddikte met 1,62 mm aan te tonen en stel dat de standaarddeviatie van de huiddikte in de groepen inderdaad 0,9 mm bedraagt. De relevante effectgrootte op de X-as in afb. 1 is dan $1,62 / 0,9 = 1,8$. Om dat verschil aan te tonen zouden er voor de eerste optie 28 dieren nodig en voor de tweede optie 22 dieren. Dat scheelt dus zes dieren.

De tweede optie vergt dat de controlegroep een factor $\sqrt{3}$ groter is dan elk van de drie behandelgroepen. Daarom moeten er van de 22 dieren acht worden toegekend aan de controlegroep. De overige 14 dieren zouden gelijk moeten worden verdeeld over de drie doseringsgroepen. Het ligt het meest voor de hand om dan één dier meer in de studie op te nemen, zodat elk van de drie groepen vijf dieren omvat.

Afb. 1 laat zien dat de aantallen dieren dalen naarmate het aan te tonen effect groter is. Dat is logisch: een groter effect is gemakkelijker aan te tonen dan een kleiner effect. Verder verloopt de afname in trapjes doordat de resultaten van een formule naar boven worden afgerond. Als de formule bij voorbeeld 31,27 dieren voorschrijft, dan hebben we dit afgerond naar 32 dieren.



Lab & OP equipment



OP toebehooren



Anesthesie



Balansen en software



Instrumentarium
onderhoud en
maatwerk



Desinfectie



Thermocage



Laboratorium
hygiëne en
disposables



Handyclean



Fondeuses

www.bioservices.nl

Our **e**xpertise generates
your quality

BIO **X**PERT

Helpt, adviseert en ondersteunt u
op vrijwel alle aspecten van in-vivo
onderzoek.

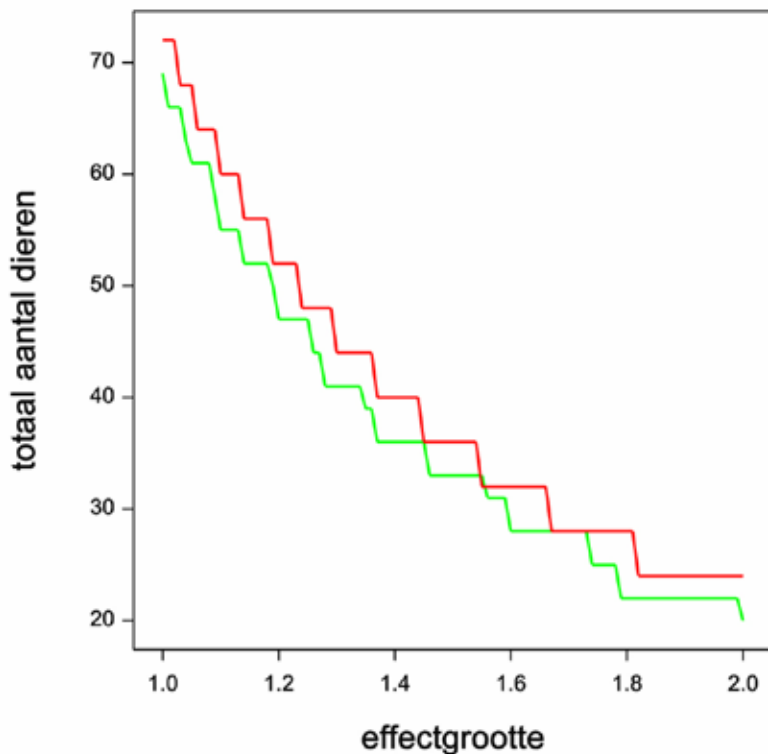
info@.BioXpert.nl



www.bioXpert.nl

In huis services

- Kolonie management
- Contract onderzoek
- Consultancy
- Cryopreservatie
- Rederivatie
- Chirurgische services
- Biologische producten
- en meer...



Afbeelding 1.
Aantal dieren om een effect aan te tonen bij gelijke en ongelijke aantallen in elke proefgroep.

Gelijke groeps grootte
Ongelijke groeps grootte

Minder dieren bij een verwachte richting van het effect

De aantallen in afb. 1 zijn berekend op basis van een éézijdige toets. Zo'n toets is zinvol als men geïnteresseerd is in een effect dat een bepaalde kant uitgaat, bijvoorbeeld een afnemende huiddikte in een behandelgroep, of juist een toename van een bloedparameter. We toetsen dan of de behandelgroep *in een bepaalde richting* afwijkt van de controlegroep.

Als we in een specifieke richting kijken om een effect te zien dan is dat effect gemakkelijker aan te tonen dan wanneer we effecten in beide richtingen voor mogelijk houden. Om aan te tonen dat huiddikten in de controlegroep 1,62 mm verschillen van huiddikten in een doseringsgroep, zonder de richting van het verschil te specificeren, zijn er voor opties 1 en 2 respectievelijk 32 en 31 dieren nodig. Vergeleken met 28 en 22 dieren bij éézijdige toetsing scheelt dat behoorlijk. Als we 10 dieren per studie kunnen besparen, dan zal dit vele dieren per jaar schelen.

Als het gaat om een rationele beslissing over groeps groottes, dan bevelen we aan om die te baseren op effecten die we willen kunnen aantonen voor de voornaamste parameters. Bij studies naar potentiële geneesmiddelen is het aannemelijk dat de voornaamste parameters van dien aard zijn dat een éézijdige toetsing op zijn plaats is. Men is immers op zoek naar een potentieel geneesmiddel en niet een middel dat de ziekte verergert.

Kleine positieve controlegroep

Zoals eerder aangegeven hoeven positieve controles bij regulatorische studies niet steeds bij elke studie meegenomen te worden, mits er voldoende ervaring is in het laboratorium waar de proef wordt uitgevoerd en het effect weinig variatie laat zien. Dit zou ook bij niet regulatorische studies naar de werkzaamheid van een potentieel geneesmiddel een zinvolle overweging kunnen zijn. Daarnaast kunnen we het aantal dieren baseren op het vermoedelijke verschil tussen de negatieve controle en de positieve controle. De positieve controle moet voor de voornaamste »

studieparameters een aanzienlijke verbetering laten zien. Het verwachte effect is dus groot. Om dit aan te tonen hoeven er dus minder dieren in de positieve controle te worden opgenomen dan in de andere behandelingsgroepen.

We illustreren dit laatste punt weer met de studie die een negatieve controlegroep en drie doseringsgroepen omvat. In dat voorbeeld berekenden we het aantal dieren op grond van de afname in huiddikte die minimaal moest worden aangetoond. De berekening resulteerde in een omvang van de negatieve controlegroep van acht dieren. Voor het vervolg van het voorbeeld gaan we uit van de verwachting dat de dieren in de positieve controlegroep een afname van 2,7 mm in huiddikte hebben in vergelijking met de dieren in de negatieve controlegroep. Met het aantal van acht dieren in de negatieve controlegroep als vertrekpunt laat een statistische poweranalyse zien dat er slechts twee dieren in de positieve controlegroep opgenomen hoeven worden om een afname van 2,7 mm aan te kunnen tonen. Een dergelijke poweranalyse (7; hoofdstuk 12) is een standaard ingrediënt in de meeste statistische pakketten.

Samenwerking

In ons voorbeeld van een niet regulatoire studie pleitten we voor vergroting van de omvang van de negatieve controlegroep ten opzichte van die van de doseringsgroepen. We zagen dat daardoor de totale studieomvang teruggebracht kan worden in vergelijking tot de optie waarin er even veel dieren in elke groep worden geplaatst. De omvang kan mogelijk verder teruggebracht worden door de gegevens van historische controlegroepen te integreren in de analyse van studieresultaten.

Tenslotte merken we op dat elke nieuwe methode op zijn merites moet worden beoordeeld. In dit geval vergt de evaluatie van de diverse maatregelen die we noemden een intensieve samenwerking van onderzoekers, statistici, opdrachtgevers en regelgevers. We zijn ervan overtuigd dat zo'n samenwerking kan helpen om het proefdiergebruik substantieel terug te dringen.

Conclusie

Vanuit ethisch, sociaal maatschappelijk, economisch en wetenschappelijk oogpunt is het zeer gewenst om te zoeken naar mogelijkheden om proefdiergebruik verder in te perken. Met dit artikel hebben we aangetoond dat er diverse opties zijn om te besparen op aantallen dieren. Bij een aantal kennisinstellingen worden bovenstaande mogelijkheden tot vermindering in toenemende mate met succes toegepast.

Literatuur

1. Dunnett CW (1955). *A multiple comparison procedure for comparing several treatments with a control*. Journal of the American Statistical Association 50: 1096-1121
2. Dunnett CW (1964). *New tables for multiple comparisons with a control*. Biometrics 20: 482-491
3. Konigsson K, Robinson S, Harlemann H (2010). *Reduction in dog numbers in regulatory one-month toxicology studies without compromising scientific quality: Challenging the status quo*. Toxicology 278: 341
4. Pfuhler S, Kirkland D, Kasper P et al. (2009). *Reduction of use of animals in regulatory genotoxicity testing: Identification and implementation opportunities – Report from ECVAM workshop*. Mutation Research/Genetic Toxicology and Environmental Mutagenesis 680: 31-42
5. Robinson S, Konigsson K, Harlemann H (2012). *Reducing dog use in one month toxicology studies without affecting scientific quality*. Toxicology Letters 211S: S180-S181.
6. Russell WMS, Burch RL (1959). *The principles of humane experimental technique* (London: Methuen)
7. Sokal RR, Rohlf FJ (2012). *Biometry* (4th edition; New York: WH Freeman)

«