INSPANNINGS-FYSIOLOGIE

In de (top)sportpraktijk worden sporters regelmatig aan testen onderworpen voor het monitoren van vooruitgang, het identificeren van zwakheden, etc. Omdat de verschillen in de topsport vaak klein zijn is het van belang dat deze testen zo nauwkeurig en betrouwbaar mogelijk zijn, zodat deze verschillen ook gemeten kunnen worden. Het gebruik van een paar simpele statistische methoden kan het interpreteren van testgegevens vereenvoudigen

Betrouwbare inspanningstesten
Simpele statistiek voor interpreteerbare resultaten

Albert Smit

Een eenmalige test zegt niet zoveel over het trainingsprogramma van een sporter. Pas als een inspanningstest voor de tweede keer wordt uitgevoerd kan een gevonden verschil mogelijk toegewezen worden aan het trainingsprogramma. Bij het interpreteren van de testresultaten moet men zich drie vragen stellen[1]:

1. Hoe goed is de geselecteerde test bij het herkennen van veranderingen bij zeer getrainde sporters?
2. Is de waargenomen verandering in conditie een "echte" verandering of simpelweg het gevolg van een "fout" of "ruis" in de test?
3. Is de grootte van de geobserveerde (en werkelijke) verandering in conditie groter dan een vooraf bepaalde drempel voor een "verandering die de moeite waard is" of voor verbetering van de testprestatie?

Worden deze vragen positief beantwoord, dan kan een gevonden verschil in testresultaten toegewezen worden aan het trainingsprogramma en bewijst dit de effectiviteit van de trainingen.

1. Hoe goed is de test?

De voorwaarden waaraan een test zou moeten voldoen wanneer deze geselecteerd wordt voor het testen van specifieke eigenschappen in een bepaalde sport zouden bekend moeten zijn bij iedereen die zich met het testen van sporters bezig houdt, maar worden hieronder nog eens genoemd:[2]

Relevantie

Om valide resultaten te behalen is het van belang dat sporters positief reageren op inspanningstesten. Hiervoor is het zeer belangrijk dat de sporters direct de relevantie van een specifieke test voor hun sport herkennen. De bekende energievereisten van de sport zouden moeten terugkomen in de test.

Specificiteit

De prestatiecapaciteiten van spiergroepen en spiervezeltypen, benodigd bij de beoefening van de sport, moeten in

de testen worden aangesproken. Spiergroepen die betrokken zijn bij de activiteit moeten ook de patronen en snelheden van de sportbeweging volgen. Fiets-, kayak-, zwem- en roeiergometers zouden gebruikt moeten worden in de specifieke sporten waarvoor ze zijn ontwikkeld wanneer het explosief en het duurvermogen gemeten worden. Dat geldt ook voor een loopband voor loopnummers. Veldtesten, waarbij de sporter zich vrij kan bewegen en niet wordt gehinderd door testapparatuur, zijn natuurlijk het meest specifiek, mits omgevingsfactoren als wind, getijde, stroming of temperatuur niet te extreem zijn.

Validiteit

Een geschikte conditietest moet daadwerkelijk dat meten wat men beoogd of claimt te meten. Met andere woorden: de test moet valide zijn. De mate van validiteit kan geschat worden door een nauwkeurige inspectie van de inhoud.

Nauwkeurigheid

Een geschikte conditietest zou ook nauwkeurig moeten zijn in vergelijking met een criteriummethode. Voorbeeld: de hoge correlatie die er is tussen de prestatie op een 20-meter shuttle run test en het maximale aërobe vermogen (zoals gemeten in een laboratorium) stelt zeker, dat de score op een 20-meter shuttle run een acceptabele maat is voor aëroob loopvermogen. Dat wil zeggen dat de nauwkeurigheid van de 20-meter shuttle run als een "veld"-maat van het duurvermogen bij lopen is bevestigd.

Praktische haalbaarheid

De vraag of een test ook praktisch haalbaar is, is ook een belangrijke overweging wanneer je kiest welke testen worden uitgevoerd. Met factoren als de beschikbaarheid van de sporters als de locatie waar de test wordt uitgevoerd moet rekening worden gehouden, wanneer de geschiktheid van een bepaalde testbatterij vastgesteld wordt.

2. Is de waargenomen verandering "echt" of het gevolg van "ruis"?

Als eenmaal een test is geselecteerd, dan moet deze worden uitgevoerd op een zodanige manier dat – ook nog op een ethisch acceptabele manier – valide en nuttige informatie wordt verkregen. De test moet betrouwbare en interpreteerbare resultaten geven. Hierbij moeten de coaches en sporters zich realiseren dat resultaten voor een groot deel beïnvloed worden door de omstandigheden waaronder de test wordt uitgevoerd. In een laboratorium kan de invloed van veel variabelen, die normaal gesproken de prestatie tijdens een wedstrijd beïnvloeden, worden geminimaliseerd. Ook hier kunnen echter kleine veranderingen in testresultaten voorkomen als gevolg van dag-tot-dag biologische variabiliteit en kleine variaties in kalibratie en het gebruik van meetapparatuur. Het Australian Institute of Sport heeft protocollen opgesteld [3] waarmee deze variaties zoveel mogelijk kunnen worden teruggebracht. Hierin staan richtlijnen om dezelfde warming-up, dezelfde volgorde van testen, dezelfde herstelperiode tussen de testen en zoveel mogelijk dezelfde klimatologische omstandigheden (temperatuur, luchtvochtigheid en luchtstroming) te creëren. De sporter zelf moet voldoende hebben gerust en niet geblesseerd of ziek zijn. Hij/zij zou op hetzelfde tijdstip op de dag getest moeten worden en een vergelijkbare vocht- en voedinginname hebben gehad. Daarna ligt de verantwoordelijkheid bij de sporters om, na zorgvuldige instructie van het testpersoneel, zichzelf bij de conditietest op een zodanige manier te presenteren dat ze in staat zijn om een prestatie te leveren die zijn of haar piek-fysiologische status van dát moment weergeeft. [2] Als dit niet het geval is, dan zouden veranderingen in conditiescores toegeschreven kunnen worden aan factoren die niets te maken hebben met de voorgeschreven trainingsmethode.

Nauwgezette controle van instrumenten, zoals kalibratie van ergometers en meetapparatuur, van sporters voor, tijdens en na de test en ook van het testpersoneel kunnen 'systematische' fouten minimaliseren, hoewel deze niet volledig zijn uit te sluiten. Kleine fouten kunnen ontstaan door de manier waarop een testcentrum met apparatuur en sporters omgaat. Met het gebruik van eenvoudige statistieken kunnen deze fouten worden gemeten én gebruikt om te bepalen of een gemeten verschil in een test toe te wijzen is aan de training, aan de slechte voorbereiding van de sporter of aan slechte kalibratie-procedures. [2] Centraal hierin staan precisie en betrouwbaarheid. Hiermee wordt elke interpretatie van de resultaten die gerealiseerd kunnen worden ondersteund. Onderstaand worden de drie begrippen die verband houden met de kwaliteit van een test toegelicht:

Precisie

Als meerdere testen uitgevoerd worden op twee of meer afzonderlijke momenten, dan zou het mogelijk moeten zijn om dezelfde resultaten te verkrijgen, ware het niet dat normaal gespro-

Bepalen van de technische fout van metingen (TEM)

De technische fout van metingen is gelijk aan de "standard error of measurement" (SEM) welke gedefinieerd wordt als: $SEM = \frac{\sigma}{\sqrt{n}}$ waarbij σ de standaarddeviatie is en *n* het aantal metingen. σ wordt berekend door

$\sigma = \sqrt{\frac{\Sigma di^2}{n}}$ waarbij *d* het verschil tussen twee metingen is. Dit maakt dat de absolute TEM berekend kan worden door

$TEM = \sqrt{\frac{\Sigma di^2}{2n}}$ en het relatieve %TEM, welke de fout als het percentage van de gemiddeld originele paren meet, door

$\%TEM = \left(\frac{TEM}{\left(\frac{M1+M2}{2}\right)} \right) x100$ M1 is het gemiddelde van de eerste serie metingen, M2 is het gemiddelde van de tweede serie metingen.

Zie tabel 1 voor een voorbeeld. Deze laat zien dat een enkele meting van deze test met deze ergometer plus of min 10,09W voor tweederde van de tijd zal zijn. De relatieve TEM is net iets meer dan 0,5% en dat is OK voor piekvermogenmetingen. Deze opstelling is dus redelijk goed in het meten van piekvermogen.

Renner	Serie X1 (W)	Serie X2 (W)	X1-X2	d^2
1	1450	1461	-11	121
2	1610	1595	15	225
3	1759	1740	19	361
4	1399	1411	-12	144
5	1755	1766	-11	121
6	1633	1620	13	169
7	1900	1893	7	49
8	1678	1698	-20	400
9	1666	1645	21	441
10	1776	1778	-2	4
Gemiddelde	1662,6	1660,7	Som d^2	2035
$TEM = \sqrt{\frac{2035}{20}} = 10,09W$				
$\%TEM = (10,09/([1662,6+1660,7]/2))x100 = 0,61\%$				

Tabel 1. Voorbeeld voor het berekenen van de TEM met een sprinttest van 6 sec. om het piekvermogen te bepalen bij 10 wielrenners op een ergometer.

ken elke herhaalde meting in meer of mindere mate vrij onvoorspelbaar zal variëren. De grootte van die variabiliteit bepaalt de mate van precisie en is karakteristiek voor een bepaalde testleider die gebruik maakt van een specifieke procedure op een specifieke variabele. Een kleine variabiliteit in opeenvolgende metingen betekent een grote precisie. Een statistische procedure die Technical Error of Measurement (TEM) genoemd wordt kan gebruikt worden om de precisie van een meting te kwantificeren. [4]

Betrouwbaarheid

Een andere eigenschap van een meting is gebaseerd op het concept van herhaalde metingen bij dezelfde personen: *betrouwbaarheid*. Betrouwbaarheid is de consistentie van opeenvolgende metingen. Deze is afhankelijk van de variabiliteit tussen sporters onderling én de variabiliteit binnen de sporter. Dit concept geeft extra informatie voor de TEM.

Interpreteerbare resultaten

Elke test zou voor de uitvoering moeten worden toegelicht aan de sporter. Er is een veel grotere kans dat de sporter maximale inspanning levert als hij of zij de redenen voor de test begrijpt, alsmede de relevantie voor zijn sport en de fysieke toewijding die het vereist. De testresultaten zouden direct moeten worden teruggekoppeld op een zodanige manier, dat de coach en sporter ze makkelijk begrijpen. In een resultatenoverzicht is het nuttig om de individuele score van elke test, het groepsgemiddelde en de positie van het individu in de groep op te nemen. De precisie (TEM) van de meting zou aan de testresultaten moeten worden toegevoegd, zodat de kans van een wérkelijke fysiologische verandering onderscheiden kan worden van een combinatie van biologische variatie en technische meetfouten.

3. Is de verandering groter dan de "kleinste verandering die de moeite waard is"?

Bij elke inspanning die geleverd wordt is sprake van een mate van variabiliteit. Een inspanningstest die herhaald wordt zal een net iets andere waarde opleveren dan de eerste test en een wedstrijd die opnieuw gelopen wordt zal in een net iets andere tijd resulteren dan de eerste wedstrijd. Voor deze variabiliteit kun je de coëfficiënt van variatie (CV) gebruiken. [5] Dit is

de standaarddeviatie uitgedrukt als percentage van het gemiddelde. Een CV van 1% bijvoorbeeld betekent dat de prestatie van een sporter van race-naar-race gewoonlijk varieert met 1m per 100m, 0.1s per 10s, 1s per 100s, etc. (zie tabel 2). De kleinste verandering (KV) in prestatie die nog de moeite waard is, is voor de meeste topsporters die op individuele onderdelen uitkomen ongeveer 0,5-1%, afhankelijk van de sport. Zie Paton en Hopkins [6] voor de methode om die kleinste verandering (KV) te bepalen. Deze waarden kunnen door een sporter gebruikt worden om zich te richten op die dingen die belangrijk zijn. Bijvoorbeeld een verbetering van 4% om bij de besten te horen of 2% per jaar gedurende drie jaar ter kwalificatie voor de Spelen.

Wanneer je een test gedaan hebt bij topsporters wil je minimaal een verschil van 0,5 x CV hebben om zeker te weten dat er vooruitgang geboekt is. Veranderingen in prestatie in labtesten zijn vaak in andere eenheden dan die voor wedstrijdprestaties. Bijvoorbeeld, een verschil van 1% in duurvermogen (Watts) op een ergometer is gelijk aan 1% in looptijd of snelheid, maar ongeveer 0,4% in fietstijd en 0,3% in roei- en zwemtijd.[5] In teamsporten, waar geen duidelijke relatie bestaat tussen testprestatie en teamprestatie, wordt de KV gesteld op een "standaard verandering in verschil", ook bekend als "Cohen's effect size (Cohen's d)". Dit wordt gedefinieerd als het verschil in gemiddelde tussen twee gelijkwaardige groepen gedeeld door de standaarddeviatie *(Δgemiddelde/SD)*. Dit is een statistische methode die gemaakt is om het gestandaardiseerde groepsgrootte-effect te bepalen.

Wat het groepsgrootte-effect is kan het beste uitgelegd worden met een voorbeeld. Stel: je hebt geen verstand van inspanningsfysiologie. Hoeveel mensen zou je dan moeten meten om er achter te komen dat, gemiddeld genomen, de mensen met de hoogste zuurstofopname het beste presteren op een uithoudingstest? Het antwoord zit in de grootte van het verschil in gemiddelde zuurstofopname tussen goede en slechte duuratleten. Hoe groter het groepsgrootte-effect, hoe sneller het verschil te zien is. Als het verschil maar klein is heb je veel meer mensen nodig om het te kunnen waarnemen. Het kleinst mogelijke groepsgrootte-effect is 0,2. De KV wordt berekend door deze 0,2 te vermenigvuldigen met de standaarddeviatie van de test. De KV voor teamsporten is dus 0,2 x SD. Dit is gelijk aan een verschuiving van het 50e percentiel naar het 58e percentiel. [5] Bijvoorbeeld, een teamsporter doet een 30m sprint test en loopt de eerste keer 5,215s. Het gemiddelde van de groep is 5,179s en de standaarddeviatie is 0,099s. De sporter zal de tweede keer 0,2 x 0,099 = 0,020s harder moeten lopen om een vooruitgang te laten zien die de moeite waard is.

SPORT	CV
Hardlopen en hordelopen tot 1500m	0,8%
Hardlopen tot 10km en steeplechase	1,1%
Cross country	1,5% (subtop)
Halve marathon	2,5% (subtop)
Hoogspringen	1,7%
Polsstokspringen, verspringen	2,3%
Discus, speerwerpen, kogelstoten	2,5%
Mountainbiken	2,4%
Triathlon (zwemmen)	1,6%
Triathlon (fietsen, individueel)	2,3%
Triathlon (lopen)	3,6%
Zwemmen	0,8%*
Wielrennen 200m	0,7%
Wielrennen 1-40km	1,3%*
* CV voor tijd. Vermenigvuldig met ~2-3 om een CV voor gemiddeld vermogen te krijgen	

Tabel 2. CV voor verschillende sporten uit gepubliceerde en ongepubliceerde studies. Vermenigvuldig met 0,5 om de KV voor topsporters te krijgen

Elke meting geeft ruis en deze kun je stellen als ± de TEM. Dit is je onzekerheid in het verschil dat je gemeten hebt. Om zeker te zijn dat je werkelijk een verandering hebt gemeten die de moeite waard is, moet de TEM kleiner zijn dan de KV (zie tabel 3). Als de TEM groter is dan de KV, moet je een betere test zoeken of de test meerdere keren herhalen met dezelfde sporter en de resultaten middelen om de ruis te verminderen (viermaal testen halveert de ruis). [5]

Interpreteren van veranderingen

Als je veranderingen vindt in testresultaten na een herhaalde test, hoe interpreteer je deze veranderingen dan voor de coach en sporter? Met andere woorden, wanneer geef je aan dat het gevonden verschil te wijten is aan de training en dat het niet aan andere factoren ligt? Hopkins [5] doet dit op de volgende manier, waarbij je stelling moet nemen over de grootte van de verandering aan de fand van de KV en rekening moet houden met de ruis (TEM):

1. Gebruik de kans dat de werkelijke waarde groter is dan de KV. Bijvoorbeeld: een sporter is met +1,5% veranderd sinds de laatste test. De ruis (TEM) is 1,0%, de KV is 0,5%. Hieruit kan berekend worden dat de kansen 76% zijn voor een positieve verandering, 16% voor een marginale verandering en 8% voor een negatieve verandering. Deze manier is exact, maar onpraktisch. Je hebt een spreadsheet nodig voor de kansberekeningen. Deze spreadsheet is te vinden op de site van Will Hopkins. [7] Stel: een wielrenner wordt getest met een maximaaltest. De eerste keer haalt hij 420W en de tweede keer haalt hij 425W. Uit een validatiestudie die je gedaan hebt is gebleken dat de TEM voor deze test 1,3% is en de KV voor deze test 1%. Voeren we dit in het Excel-bestand in, dan zien we dat er 54% kans is dat dit een positief verschil is, 34% kans dat dit een triviaal verschil is en 12% kans dat dit een negatief verschil is. De testresultaten neigen naar een werkelijk verschil, maar het is niet overduidelijk.
2. Gebruik waarschijnlijke limieten voor de werkelijke waarde. De gemakkelijkste waarschijnlijke limieten zijn de geobserveerde verandering ± de TEM. De werkelijke waarde zou tussen deze twee limieten kunnen zitten. Dat wil zeggen, "een kans van 50%" of "kans van 1:1" of "mogelijk". Interpreteer de limieten als positief, marginaal of negatief. Noem het effect "duidelijk" als beide limieten hetzelfde zijn (beide positief of beide negatief). Duidelijke gevallen hebben in 76% van de gevallen gelijk. Bijvoorbeeld: een sporter is met 2,5% veranderd sinds de laatste test, de KV is 1,0%. Als de TEM ± 2,0% is, is het verschil onduidelijk. Als de TEM ± 1,0% is, is het verschil duidelijk. Neem hetzelfde voorbeeld van de wielrenner. De eerste keer trapt hij 420W. De KV is 1%. Hij moet dus in ieder geval 420W + 1% x 420W = 424,2W rijden om duidelijk beter zijn dan de vorige keer. De waarschijnlijke limieten zijn de geobserveerde verandering ± de TEM, dus 425 ± 1,3% x 420 = 425 ± 5,46W. De werkelijke waarde van de tweede test ligt dus binnen de limieten 419,5W en 430,5W. De 424,4W van de eerste test vermeerderd met de KV ligt binnen deze limieten en er heeft dus een marginaal verschil plaatsgevonden. Als de renner 430W getrapt had, dan zouden de limieten van 424,5 tot 435,5 lopen en dan zou er wel een duidelijk positief verschil zijn.
3. Gebruik deze simpele regels:
   a. Als de test goed is (ruis kleiner dan de KV), interpreteer dan alle veranderingen als duidelijk positief, triviaal of negatief. Je zult in 50% van de tijd gelijk hebben (vaak nog meer).
   b. Als de test slecht is (ruis groter dan de KV), interpreteer de verschillen dan alleen wanneer deze groter zijn dan de ruis. Dat wil zeggen, elke verandering groter dan de ruis is positief (of negatief).
   c. Elke verandering < ruis is onduidelijk.

   Uitspraken over positief of negatief verschil gelden in meer dan 50% van de gevallen. Nogmaals het voorbeeld van de wielrenner: de huidige situatie wordt als "slecht" omschreven, omdat de ruis (TEM = 1,3% = 5,5W) groter is dan de KV (1,0% = 4,2W). Wanneer 425W in de tweede test getrapt wordt is dit minder dan de TEM en is de verandering dus triviaal. Wanneer 430W getrapt wordt, is dit groter dan de ruis en is dit een duidelijk positieve verandering. Wanneer we de testopstelling zodanig kunnen aanpassen dat de ruis kleiner dan 1,0% wordt, dan kunnen we van elke verandering groter dan de KV zeggen dat dit een duidelijk positieve of negatieve verandering is.
4. Geef de ruis de 'schuld' voor extreme testresultaten. Wanneer je

Test	TEM		KV		Waardering
	Absoluut	%	Absoluut	%	
Lichaamsgewicht (kg)	1.5	2	1.4	1.6	OK
Lengte (cm)	1.0	0.5	2	1.2	Goed
Huidplooien (mm)	1.5	3	2.7	4.4	Goed
AFL agility (s)	0.13	1.5	0.05	0.6	Marginaal
20m sprint (s)	0.04	1.5	0.02	0.8	Marginaal
20m shuttle (lvl)	0.04	3.5	0.03	2.4	Marginaal
Verticale sprong (cm)	1.3	2	1.3	2	OK
Sit & Reach (cm)	1.2	10	1.5	12	OK
Herhaalde sprints (s)	0.36	1.3	0.3	1.0	OK

Tabel 3. Waarden voor de TEM en kleinste verandering die de moeite waard is (KV) van algemene conditietesten, geproduceerd door het Department of Physiology, Australian Institute of Sport. Testen waar de KV veel groter is dan de TEM worden gewaardeerd als "Goed", die met gelijke TEM en KV scoren als "OK" en die met een grote TEM en een kleine KV scoren "Marginaal" (uit: Pyne, 2003).

extreme resultaten vind die je niet kunt verklaren door de trainingen die gedaan zijn, dan kun je de ruis de schuld geven van de extreme bevindingen en zeggen dat de werkelijke waarde waarschijnlijk ergens in het midden ligt.

De testresultaten kunnen nu aan de sporter en coach teruggekoppeld worden op een zodanige manier, dat het duidelijk en begrijpelijk is voor coach en sporter. Daarbij is tevens aangegeven of het gevonden verschil met de vorige test duidelijk positief of negatief, danwel marginaal is.

Samenvatting

Elke (inspannings)test zou aan de voorwaarden van relevantie, specificiteit, praktische haalbaarheid, validiteit en nauwkeurigheid moeten voldoen. Van elke test zou de precisie (TEM) en betrouwbaarheid bekend moeten zijn en elke test zou interpreteerbare resultaten moeten opleveren. Testresultaten zouden bij het interpreteren getoetst moeten worden aan een vooraf opgestelde drempel (de kleinste verandering die de moeite waard is), waarbij rekening gehouden wordt met de ruis. Resultaten kunnen dan duidelijk interpreteerbaar aan sporter en coach worden teruggekoppeld.

Referenties

1. Pyne, D.P. (2003). Interpreting the results of fitness testing. Speakers notes of the Gastrolyte International Science and Football Symposium. Melbourne, Victorian Institute of Sport, (URL: http://www.vis.org.au/downloads/science/Pyne.pdf).

2. Pyke, F. (2000). Introduction. In: Gore, C.J. (Ed).Physiological tests for elite athletes (pp xii-xiv). Champaign, Illinois: Human Kinetics.

3. Gore, C.J. (2000) (Ed).Physiological tests for elite athletes. Champaign, Illinois: Human Kinetics.

4. Pederson, D.G. & C.J. Gore (1996). Antropometry measurement error. In: Norton, K.I. & T. Olds (Eds.). Anthropometrica (pp 77-96). Sydney, University of New South Wales Press.

5. Hopkins, W.G. (2004). How to interpret changes in an athletic performance test. Sportscience 8, 1-7, (URL: http://www.sportsci.org/jour/04/wghtests.htm).

6. Paton, C.D. & W.G. Hopkins (2005). Competitive performance of elite olympic-distance triathletes: reliability and smallest worthwhile enhancement. Sportscience 9, 1-5, (URL: http://www.sportsci.org/jour/05/wghtri.htm).

7. Hopkins, W.G. (2001). Spreadsheet for assessing an individual. A new view of statistics. (URL: http://sportsci.org/resource/stats/).

Over de auteur

Albert Smit heeft Bewegingswetenschappen gestudeerd en is sinds 2003 werkzaam bij NOC*NSF als wetenschappelijk medewerker. Vanuit NOC*NSF is hij expert-adviseur voor de KNWU en onderdeel van de begeleidingsstaf van de nationale selectie baanwielrennen. Hierbij is hij verantwoordelijk voor het uitvoeren van testen en het maken van trainings- en wedstrijdanalyses.
