

Arnold Bakker, Toon Taris en Jan de Jonge

## 10.1 Inleiding

De afgelopen decennia zijn er talloze studies verricht naar de relatie tussen arbeid en gezondheid. Hoewel dit onderzoek belangrijke inzichten heeft opgeleverd, kleven aan veel studies methodische problemen die de betrouwbaarheid en validiteit van de conclusies bedreigen. Zo is het bijvoorbeeld onmogelijk om uitspraken te doen over causaliteit op basis van cross-sectioneel onderzoek, en worden experimentele studies soms gekenmerkt door een gebrekkige generaliseerbaarheid naar het werkelijke arbeidsleven. In dit hoofdstuk wordt daarom aandacht besteed aan belangrijke overwegingen bij onderzoek op het gebied van de A&G-psychologie.

Dit hoofdstuk bestaat uit drie delen. In het eerste deel zullen drie veelvoorkomende typen onderzoeksontwerpen ('designs') de revue passeren. Er wordt achtereenvolgens ingegaan op cross-sectioneel onderzoek (par. 10.2), longitudinaal onderzoek (par. 10.3) en experimenteel onderzoek (par. 10.4). Bij de bespreking van deze designs en in de rest van het hoofdstuk zal het callcenter uit box 10.1 steeds als voorbeeld worden gebruikt. In het tweede deel wordt ingegaan op methodische aspecten die een rol spelen in veel A&G-psychologisch onderzoek. In paragraaf 10.5 staat het onderscheid tussen objectieve en subjectieve benaderingen centraal. Besproken worden de theoretische, methodische en praktische redenen om op zoek te gaan naar objectieve indicatoren van de werksituatie, evenals de methodische problemen bij het verrichten van objectieve metingen.

In paragraaf 10.6 is er aandacht voor de invloed van een derde variabele Z op de relatie tussen twee variabelen X en Y. Achtereenvolgens worden drie van deze invloeden besproken, die respectievelijk bekend staan als het 'derde variabeleprobleem', mediatoreffecten en moderatoreffecten. Tot slot wordt in het derde deel van dit hoofdstuk aandacht besteed aan twee technieken die tot nu toe relatief weinig worden gebruikt in onderzoek maar wel zeer waardevol kunnen zijn: meta-analyse (par. 10.7) en multiniveau-analyse (par.

10.8). Deze analysetechnieken zullen naar verwachting in toekomstig A&G-psychologisch onderzoek in een behoefte voorzien.

### **Box 10.1 Objectieve en subjectieve stressoren en mediator- en moderatoreffecten in een callcenter**

Call Us is het callcenter van een jong telecombedrijf met ongeveer zeshonderd medewerkers. Het callcenter bestaat nu vier jaar en is opgebouwd uit een frontoffice en een backoffice. In de frontoffice zijn de teleadvisers in teamverband werkzaam. Zij nemen de binnenkomende telefoontjes aan, en lossen eenvoudige klachten en problemen op. Zij fungeren vooral als filter. De klachten worden door de teleadvisers beschreven in een zogenoemde inbox op de computer en per e-mail doorgestuurd naar de backoffice.

In de backoffice werken de teleconsultants. Zij doen het uitzoekwerk, lossen problemen op en bellen klanten hierover terug. Het telecombedrijf is de afgelopen jaren enorm gegroeid. De toenemende verkoop van mobiele telefoons heeft ertoe geleid dat de omzet en winst zijn gestegen, maar ook dat er een toename is van het aantal klachten over rekeningen, storingen van telefoongesprekken en mobiele telefoons die niet werken.

Hoewel ook het callcenter is meegegroeid in aantal medewerkers, is het ziekteverzuim en personeelsverloop bij Call Us disproportioneel hoog: ongeveer 20% van de medewerkers is ziek, en het afgelopen jaar vertrok 15% van het personeel. Gesprekken met veertig teleadvisers geven een beeld van de oorzaken van deze ongunstige cijfers. Teleadviser Janssen geeft aan: 'Soms zijn er pieken. Het kan echt stressvol zijn door de onderbezetting en verkeerde planning. Als je voortdurend bezig bent word je daar ontzettend moe van.' Zijn collega Peters verwijst naar een andere objectieve stressor: 'Het is soms erg moeilijk om de norm van 47 seconden per telefoongesprek te halen. Doordat je klantgericht wilt werken, en doordat de computersystemen niet altijd goed functioneren, duurt het soms langer.'

Teleadvisers hebben vaak te maken met boze klanten. Meneer Dirksen: 'Het is wel moeilijk dat je steeds met klagende mensen te maken hebt. Je wordt er moe van en het geloof in wat je doet raak je kwijt. De irritatiegrens komt lager te liggen. Je krijgt ook veel te weinig informatie over positieve resultaten. Daardoor krijg je de indruk dat alles altijd slecht gaat' (subjectieve interpretatie van stressoren). Zijn collega Simons zegt: 'Het gros van de klanten is boos of agressief. In trainingen is nooit uitgelegd hoe je dat moet aanpakken. Het gevolg is dat je afgestompt en uitgeput raakt. Ik zie om mij heen dat collega's het soms niet meer trekken en ziek thuisblijven' (mediatoreffect van burnout). Mevrouw Roeten beaamt: 'Het gaat er soms vrij stevig aan toe. Dan moet je even je headset neerleggen en een glas water drinken. Je krijgt veel agressie over je heen. Dan is het beter om even rust te nemen en bij collega's je verhaal te doen. Dat zou ik twee jaar geleden niet hebben gedaan' (moderatoreffect van sociale steun).

## 10.2 Cross-sectioneel onderzoek

Wanneer onderzoekers vaststellen dat werknemers die zich langdurend of intensief inspannen over het algemeen vermoeider zijn, geven ze betekenis aan een belangrijk aspect van vermoeidheid. Wanneer voorts wordt vastgesteld dat vermoeide werknemers meer fouten maken en vaker verzuimen, begint men te begrijpen wat vermoeidheid inhoudt. Men leert aldus over de relaties die betekenis geven aan het begrip ‘vermoeidheid’. Verreweg het meeste onderzoek op het terrein van de A&G-psychologie richt zich op dit soort relaties in cross-sectioneel- of dwarsdoorsnedeonderzoek (Taris, De Jonge & Ybema, 2005).

Dit betekent dat er op een moment in de tijd metingen worden verricht bij een groep van werknemers, bijvoorbeeld metingen van werkdruk, emotionele taakeisen en vermoeidheid met een vragenlijst. Zulke gegevens kunnen vervolgens met elkaar in verband worden gebracht. Er is sprake van een positieve correlatie tussen bijvoorbeeld emotionele taakeisen en vermoeidheid bij callcentermedewerkers (zie box 10.1), wanneer een toename van het aantal klachten over rekeningen en storingen van telefoongesprekken gepaard gaat met een toename van vermoeidheidsverschijnselen.

Voorts kan men met cross-sectioneel onderzoek vaststellen wat de prevalentie is van bepaalde aspecten van arbeid en gezondheid (zie ook hfst. 12). Zo kan men nagaan hoe hoog de emotionele taakeisen in een callcenter zijn en wat de prevalentie is van burnout. Met behulp van referentiegegevens van andere callcenters (‘benchmarks’; zie hfst. 6 en 11) kan men nagaan of de emotionele taakeisen relatief hoog zijn en of veel medewerkers last hebben van burnout. Op basis van deze gegevens kunnen onderzoekers bijvoorbeeld een theorie vormen over de oorzaken van het burnoutsyndroom, of aanbevelingen doen over de manier waarop een organisatie burnout onder haar medewerkers kan terugbrengen of voorkomen (zie hfst. 17).

Een belangrijk probleem van eenmalige metingen is dat oorzaak en gevolg moeilijk – zo niet onmogelijk – te scheiden zijn, en er geen harde uitspraken kunnen worden gedaan over causaliteit. Onderzoekers willen echter vaak weten wat de oorzaken zijn van bepaalde verschijnselen (bijv. vermoeidheid of burnout). Is werkdruk inderdaad de belangrijkste veroorzaker van chronische vermoeidheid? Leidt confrontatie met lastige patiënten daadwerkelijk tot burnout? In dergelijke gevallen wil men dus uitspraken doen over de oorzakelijke relatie tussen werkkenmerken en gezondheid. Een probleem van cross-sectioneel onderzoek is dat we een ‘foto’ maken van de werksituatie, terwijl het normale arbeidsleven beter is weer te geven als een ‘film’. Met een film wordt immers duidelijk wat de ordening van fenomenen is in de tijd.

Zo is intussen uit onderzoek bekend dat werkdruk niet alleen kan leiden tot stressreacties, maar dat stressreacties ook weer van invloed kunnen zijn op (de waarneming van) werkdruk. Dit betekent dat vermoeidheid bij callcentermedewerkers in de loop der tijd kan leiden tot een toename van hun werkdruk (bijv. doordat ze meer fouten maken), en dat huisartsen met een negatieve, cynische werkhouding in toekomstige interacties vaker te maken

kunnen krijgen met lastige of zelfs agressieve patiënten. Deze wederkerige of reciproke relaties tussen taakeisen en stressreacties worden in paragraaf 10.3 besproken.

### 10.3 Longitudinaal onderzoek

Recentelijk zien we een duidelijke toename van onderzoek op het terrein van de A&G-psychologie waarbij gebruik wordt gemaakt van meer dan één meting. Bij dit zogeheten longitudinaal onderzoek worden dezelfde werknemers in de tijd gevolgd (bijv. gedurende enkele maanden of jaren) en worden twee of meer metingen verricht op verschillende tijdstippen. Er is een redelijke consensus in de literatuur over de criteria waaraan longitudinaal onderzoek moet voldoen (zie bijv. De Lange, Kompier, De Jonge, Taris & Houtman, 2001; Taris, 2000). In dit hoofdstuk worden de drie belangrijkste criteria besproken, die respectievelijk betrekking hebben op het onderzoeksmodel, het tijdsinterval en de steekproef.

#### 10.3.1 Onderzoeksmodel

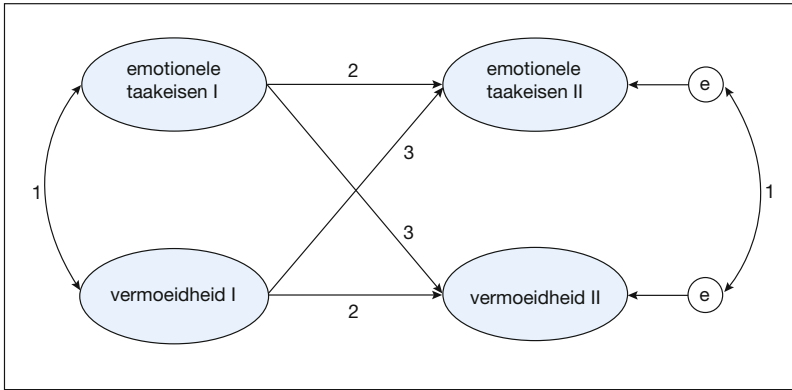
In figuur 10.1 wordt een model afgebeeld dat in de methodologie bekend staat als een compleet paneldesign voor twee metingen en twee variabelen (Zapf, Dormann & Frese, 1996). In een zogenoemde panelstudie wordt dezelfde groep deelnemers herhaaldelijk ondervraagd (in dit geval tweemaal), waarbij gebruik wordt gemaakt van dezelfde variabelen. Anders gezegd: in een compleet paneldesign worden de onafhankelijke en afhankelijke variabelen op meer tijdstippen (bijv. met een tussenliggende periode van een jaar) met dezelfde meetinstrumenten bij dezelfde personen gemeten.

In figuur 10.1 zijn verschillende soorten relaties te onderscheiden. Er is sprake van synchrone relaties tussen variabelen die op hetzelfde tijdstip zijn gemeten (de verbanden met nummer 1). Toegepast op het callcenter uit box 10.1, kan dit het verband zijn tussen emotionele taakeisen op tijdstip 1 (T1) en vermoeidheid op datzelfde tijdstip. Dit betekent dat medewerkers die vaak worden blootgesteld aan vervelende, klagende klanten, vermoeider zullen zijn dan anderen. Autocorrelaties verwijzen naar de verbanden tussen een bepaalde variabele zoals gemeten op T1 en dezelfde variabele op T2 (de relaties met nummer 2 in de figuur), zoals het verband tussen vermoeidheid op T1 en vermoeidheid op T2.

Dat verband, dat in feite de stabiliteit van een construct over tijd weergeeft, is vaak vrij sterk. Uit onderzoek bij huisartsen bleek bijvoorbeeld dat vermoeidheid een samenhang vertoonde van 0,50 met vermoeidheid zoals vijf jaar later gemeten werd (Bakker, Schaufeli, Sixma, Bosveld & Van Dierendonck, 2000). Dit betekent dat vermoeidheid een belangrijke voorspeller is van zichzelf: het gaat niet over met een nachtje goed doorslapen of na een vakantie van drie weken.

Tot slot gaat het bij gekruiste verbanden over de tijd ('cross-lagged' relaties) over de relatie tussen een bepaalde variabele zoals gemeten op T1 en een

andere variabele zoals gemeten op T2 (nummer 3). In het callcenter betreft dit bijvoorbeeld het effect van T1 emotionele taakeisen op T2 vermoeidheid. Dit verband is naar verwachting positief: hoe meer emotionele taakeisen men heeft op T1, hoe vermoeider men zich voelt op T2.



*Figuur 10.1*

Compleet paneldesign voor twee metingen van emotionele taakeisen en vermoeidheid.

Om een samenhang in een compleet paneldesign, zoals afgebeeld in figuur 10.1, in causale termen te kunnen duiden moet ten minste aan drie eisen worden voldaan (Cook & Campbell, 1979):

- 1 er is een statistisch verband tussen emotionele taakeisen (X) en vermoeidheid (Y);
- 2 de oorzakelijke variabele X gaat in de tijd vooraf aan uitkomstvariabele Y;
- 3 de invloed van 'derde variabelen' (bijv. negatieve affectiviteit, zie par. 10.6, en persoonlijkheid, zie hfst. 15) op de relatie tussen X en Y kan worden uitgesloten.

Taris (2000) voegt hieraan toe dat er ook een aannemelijke theoretische argumentatie moet zijn voor de stelling dat X de oorzakelijke variabele is die leidt tot de uitkomstvariabele Y. Cross-sectioneel onderzoek voldoet in ieder geval niet aan de tweede eis, en daarom kunnen causale relaties beter onderzocht worden met een longitudinaal onderzoeksontwerp. Er kan een onderscheid gemaakt worden tussen drie soorten causale relaties: gewone, omgekeerde en reciproke causale relaties.

Gewone causale relaties zijn relaties die in overeenstemming zijn met de onderzoekshypothese dat variabele X een effect heeft op variabele Y. Veel onderzoekers zijn van mening dat veelvuldige of langdurige blootstelling aan werkdruk op den duur leidt tot vermoeidheid. In een compleet paneldesign kan ook worden nagegaan of er sprake is van omgekeerde causaliteit (Y beïnvloedt X) en reciproke causaliteit (X en Y beïnvloeden elkaar). In het voorbeeld zou omgekeerde causaliteit betekenen dat vermoeidheid er op den duur toe leidt dat de werkdruk toeneemt.

Dit is niet zo vreemd als wellicht op het eerste gezicht lijkt. Werknemers die chronisch vermoeid zijn, zullen bijvoorbeeld gemiddeld genomen meer fouten maken bij de uitvoering van hun werk. Stel dat een vermoeide callcentermedewerker per abuis de rekeningen van klanten verwisselt. De klanten zullen dan aan de bel trekken en klagen over de dienstverlening. Kortom, in deze situatie neemt de werkdruk toe, want naast de normale werkzaamheden zal de medewerker zich moeten inspannen om alsnog de juiste rekeningen bij de juiste klanten te laten bezorgen.

Bovendien is het mogelijk dat werkomstandigheden zoals werkdruk en emotionele taakeisen worden beïnvloed door de perceptie van deze stressoren (Zapf e.a., 1996). Net als de neiging van depressieve mensen om hun omgeving negatiever waar te nemen, is het mogelijk dat vermoeide werknemers meer werkstressoren waarnemen en meer klagen over hun werkdruk. Medewerkers van een callcenter zullen het laatste uur van hun dienst over het algemeen als belastender ervaren dan het eerste uur (zie hfst. 3). Volgens de sociale-informatieverwerkingsbenadering (Salancik & Pfeffer, 1978) zullen cognities en emoties ten aanzien van het werk een rationalisatieproces in gang zetten waardoor werknemers de kenmerken van hun werk cognitief zodanig construeren dat zij consistent zijn met de sociale context.

Hoewel het onderzoek naar omgekeerde of reciproke relaties nog beperkt is, blijkt uit de literatuur dat zij wel degelijk bestaan. Er zijn bijvoorbeeld omgekeerde causale effecten gevonden tussen sociale steun van collega's en mentale gezondheid, emotionele taakeisen en cynisme, en tussen werkdruk en vermoeidheid (zie bijv. De Jonge e.a., 2001; De Lange, 2005). Over het algemeen suggereren de resultaten van deze studies dat stressreacties zowel de uitkomst als de oorzaak van werkstressoren kunnen zijn, al is het effect van stressreacties op de werkstressoren over het algemeen duidelijk zwakker dan het omgekeerde effect. Stress kan dus op termijn bijdragen aan (de waarneming van) een ongunstige werkomgeving.

Hoewel het aannemelijk lijkt dat stressoren de initiator zijn van het stressproces, waarbij bijvoorbeeld langdurige blootstelling aan klagende klanten (emotionele taakeisen) leidt tot chronische vermoeidheid bij callcentermedewerkers, is het ook denkbaar dat werknemers met vermoeidheidsverschijnselen het proces zelf in gang zetten. Suboptimaal functioneren zou in dit laatste geval op termijn meer lastige klantcontacten kunnen veroorzaken. Het behoeft weinig betoog dat men bijvoorbeeld bij reïntegratie van 'opgebrande' werknemers in het callcenter (zie hfst. 8) serieus rekening moet houden met zulke omgekeerde causale effecten, terwijl voor andere medewerkers veeleer aan gewone causale effecten moet worden gedacht. Door de parameters die horen bij de gekruiste verbanden over de tijd in figuur 10.1 (nummers 3) statistisch met elkaar te vergelijken, kan men uitrekenen welke factor in het complete paneldesign causaal dominant is.

### 10.3.2 Tijdsinterval

Een compleet paneldesign biedt geen garantie voor adequaat longitudinaal onderzoek. Van groot belang is ook het tijdsinterval tussen de metingen die

worden uitgevoerd. Er is weinig overeenstemming over de juiste lengte van de gebruikte tijdsintervallen in longitudinale studies. Zapf e.a. (1996) stellen dat de lengte van de tijdsintervallen tussen de metingen gelijk zou moeten zijn. Anderen betogen daarentegen dat sommige processen beter met ongelijke tijdsintervallen bestudeerd kunnen worden, omdat het lastig is om van tevoren precies te voorspellen na hoeveel tijd stressreacties optreden. Een principe waarover wel consensus bestaat is dat de tijdsintervallen zo opgesteld dienen te worden dat een variabele X in ieder geval de tijd heeft om variabele Y te beïnvloeden. Zo is het onwaarschijnlijk dat een stressvolle baan in de horeca binnen drie weken tot burnout zal leiden. Burnout is immers een chronische stressreactie die vaak pas tot uiting komt nadat de klachten langer dan een jaar hebben aangehouden (zie hfst. 17).

Tijdsintervallen dienen verder rekening te houden met tussentijdse veranderingen of storende factoren die de te onderzoeken relatie kunnen beïnvloeden. Voorbeelden van zulke storende factoren worden genoemd in tabel 10.1. Omdat er geen gouden regel is voor de keuze van tijdsintervallen, dienen onderzoekers de lengte van tijdsintervallen zoveel mogelijk te bepalen aan de hand van de veronderstelde ontstaansgeschiedenis van de processen die ze bestuderen (De Lange e.a., 2001). Meestal worden tijdsintervallen opgesteld op basis van de praktische mogelijkheden van een onderzoek. Een theoretische en methodologische onderbouwing van de lengte van deze tijdsintervallen is echter beter en houdt in dat de keuze voor een bepaald tijdsinterval wordt verantwoord in het licht van eerdere theorievorming en/of resultaten van eerder onderzoek.

Als we bijvoorbeeld in een studie met drie meetmomenten willen nagaan wat het effect is van cognitieve en emotionele taakeisen op ziekteverzuim, dan moeten de tijdsintervallen aansluiten bij de verzuimcijfers van de onderzochte organisatie. Hierbij geldt dat het tijdsinterval korter kan zijn naarmate het verzuimpercentage hoger is. Wanneer slechts 2% van de medewerkers op jaarbasis verzuimt, dan is de variantie in verzuimscores mogelijk zo beperkt dat het moeilijk wordt om bij een gegeven steekproefomvang binnen één jaar relaties vast te stellen met werkkenmerken. In die situatie ligt het voor de hand om een relatief lang tijdsinterval (bijv. van twee jaar) te hanteren.

Merk overigens op dat in deze laatste opzet de kans op versturende gebeurtenissen (bijv. een reorganisatie, andere collega's, ander werk) weer toeneemt. Wanneer het verzuim daarentegen rond de 20% ligt, is het bij een steekproef van voldoende omvang mogelijk om op veel kortere termijn effecten vast te stellen (bijv. reeds na drie maanden). De keuze voor een bepaald tijdsinterval hangt dus af van allerlei factoren, zoals het aantal deelnemers aan het onderzoek, de beschikbaarheid van middelen of de prevalentie van de te onderzoeken fenomenen. Het ligt voor de hand om hierbij pragmatische keuzen te maken.

Tabel 10.1 Storende factoren in longitudinaal onderzoek.	
<i>storende factor</i>	<i>omschrijving</i>
1 tussentijds voorval	Het is mogelijk dat de medewerkers in het callcenter gaandeweg vermoeder worden, niet als gevolg van confrontatie met klagende klanten maar bijvoorbeeld door gebeurtenissen in de privé-situatie (bijv. het overlijden van een dierbare).
2 Rijping of groei	Een verandering in de afhankelijke variabele (bijv. verminderde vermoeidheid) kan ook worden veroorzaakt door rijpingseffecten bij de onderzoeksgroep, bijvoorbeeld doordat de werknemers wijzer, milder en meer ervaren geworden zijn, en niet zozeer door de werkomstandigheden (afgenomen blootstelling aan klagende klanten).
3 testeffect	Als bij deelnemers aan een onderzoek meerdere metingen worden verricht, is het mogelijk dat men bij de latere metingen positiever of juist negatiever scoort dan bij de eerste meting. Dit wordt mogelijk niet zozeer veroorzaakt door de werkomstandigheden, maar doordat de vragen op T1 mensen aan het denken hebben gezet, of hebben aangezet tot bepaald gedrag.
4 instrumentatie	Stel dat in het callcenter bij een eerste meting van emotionele taakeisen en vermoeidheid mondeling een korte vragenlijst wordt afgenomen, en dat de tweede meting plaatsvindt door middel van computers. Ook al zijn de vragen die worden gesteld op T1 en T2 hetzelfde, dan vinden de voor- en nameting toch plaats onder verschillende condities. Het is in dit geval denkbaar dat verschillen die worden gevonden tussen voor- en nameting niet toe te schrijven zijn aan daadwerkelijke veranderingen, maar aan deze externe factor, instrumentatie.
5 Selectie en statistische regressie	De aanleiding voor veel onderzoek op het gebied van de A&G-psychologie is dat het ziekteverzuim of personeelsverloop relatief hoog is en dat het management van een organisatie hier graag iets aan wil doen. Het is dan ook niet denkbeeldig dat men uit de populatie juist die mensen selecteert die zeer veel last hebben van werkstress. De scores van deze subgroep wijken waarschijnlijk af van het gemiddelde van de gehele onderzoekspopulatie, mede op basis van toevallige factoren en onbetrouwbaarheid van de meetinstrumenten. De kans dat de geselecteerde groep na een jaar minder vermoeidheidsverschijnselen rapporteert, is dan groter dan dat zij evenveel vermoeidheid rapporteert. De toevalsfactoren die eerder leidden tot een hoge score op vermoeidheid zijn wellicht verdwenen, en de onbetrouwbaarheid van het meetinstrument zal er wellicht nu voor zorgen dat men wat lager scoort. Dit verschijnsel staat bekend als statistische regressie (of regressie naar het gemiddelde van de groep als geheel).



<i>storende factor</i>	<i>omschrijving</i>
6 uitval	Er is sprake van uitval of mortaliteit wanneer bij een nameting minder mensen meedoen aan het onderzoek dan bij de voormeting. Dit probleem is over het algemeen groter naarmate de tussenliggende periode langer is. Mensen die in onderzoek uitvallen, vormen vaak een 'selecte' groep. Uit eigen onderzoek bij huisartsen met een tussenliggende periode van vijf jaar bleek bijvoorbeeld dat vooral artsen die hoog scoorden op burnout hun praktijk opzegden en dus niet beschikbaar waren voor de tweede meting. Indien nu het percentage burnout op de tweede meting voor de onderzoeksgroep als totaal lager is dan op T1, dan is dat dus mogelijk te herleiden tot de factor 'uitval' en niet tot bijvoorbeeld de afname in het aantal (lastige)patiëntcontacten.
7 niet-representatieve steekproeven	Als een onderzoek bij bepaalde groepen werknemers op een bepaalde plaats, op een bepaald tijdstip en onder bepaalde omstandigheden heeft plaatsgevonden, is het de vraag of de resultaten ervan zonder meer gegeneraliseerd kunnen worden naar andere plaatsen, andere tijdstippen en andere omstandigheden. Andere callcenters zouden weleens een heel andere organisatiestructuur en andere werkkenmerken kunnen hebben. Als generalisatie wel mogelijk is, dan zijn de resultaten extern valide.

Bron: Cook en Campbell (1979).

### 10.3.3 Steekproef

Voor een goed begrip van de resultaten van longitudinaal onderzoek is het van belang om de onderzochte steekproef goed te beschrijven. Dit houdt in dat niet alleen de steekproefomvang, maar ook de homogeniteit van de groep wordt beschreven. Het is lastig om in algemene termen aan te geven wat een adequate steekproefgrootte is, want dat hangt onder meer af van de verwachte sterkte van de verbanden tussen variabelen (effectgrootte). Gezien de voor A&G-psychologisch (vragenlijst)onderzoek typerende effectgroottes is het aanbevelenswaardig te streven naar minimaal enkele honderden deelnemers.

Bij onderzoek zal in de regel niet iedereen die tot de beoogde steekproef behoorde daadwerkelijk meewerken aan het onderzoek. Sommige mensen hebben geen tijd, anderen geen zin en weer anderen kunnen helemaal niet worden bereikt (bijv. doordat de verstuurde vragenlijst hen niet bereikt vanwege een verhuizing). Het is van belang om te onderzoeken in hoeverre door deze non-respons bepaalde selectie-effecten zijn opgetreden. Het is bijvoorbeeld mogelijk dat de mensen die niet aan het onderzoek meedoen vanwege verhuizing heel andere mensen zijn dan de anderen (bijv. doordat ze jonger zijn dan diegenen die wel meededen, ze nog geen kinderen hebben of hun kinderen al het huis uit zijn, ze minder gebonden aan hun buurt zijn). Mensen die geen zin hebben om mee te werken, vinden het onderwerp misschien minder interessant omdat het voor hen minder relevant is (bijv.

omdat ze helemaal geen klachten hebben over hun werk, of juist omdat ze toch al besloten hebben een andere baan te zoeken). Dat roept weer de vraag op of diegenen die wél aan het onderzoek meedoen representatief zijn voor de beoogde populatie.

Het is daarom belangrijk om het effect van deze mogelijke selectieve non-respons op de resultaten van het onderzoek in kaart te brengen. In cross-sectioneel onderzoek kan men bijvoorbeeld bekijken of de steekproef lijkt op de beoogde populatie in termen van opleiding, leeftijd en sekse. In een longitudinaal onderzoek met drie metingen zou men kunnen nagaan of werknemers die wel meedoen op T1, maar niet op T2 en T3, op T1 hoger scoren op vermoeidheid dan werknemers die aan alle drie de metingen meedoen. Als dit het geval is, dan is er sprake van een selectie-effect: het zogenoemde ‘healthy worker’-effect. Soortgelijke problemen kunnen optreden wanneer een groep werknemers wel deelneemt aan het onderzoek op T2, maar niet op T1 (bijv. omdat zij toen op vakantie waren; zie ook tabel 10.1).

Het grote probleem met selectiviteit is uiteraard dat de onderzoeksresultaten niet representatief zijn voor de totale populatie van werknemers (in een organisatie of in de beroepsbevolking) en dus niet zonder meer gegeneraliseerd kunnen worden. Een ander probleem is dat er bij selectie veelal sprake is van ‘restriction of range’-effecten. Toegepast op de callcentermedewerkers uit box 10.1 betekent dit dat de variantie in scores op vermoeidheid beperkt kan zijn, doordat vooral de gezonde en vitale medewerkers gehoor geven aan de oproep om deel te nemen aan een tweede onderzoek of doordat alleen de mensen met veel werkstress geïnteresseerd zijn in deelname om zo hun positie nogmaals bij de onderzoekers (en bij de opdrachtgevers) onder de aandacht te brengen. De beperking van de spreiding in de antwoorden op onderzoeksvragen maakt dat eventuele verbanden tussen bijvoorbeeld werkdruk en vermoeidheid moeilijker zijn vast te stellen.

#### 10.4 Experimenteel onderzoek

Een experiment is een wetenschappelijk onderzoek waarbij een onderzoeker onder gecontroleerde omstandigheden ten minste één onafhankelijke variabele manipuleert (bijv. werkdruk) en nagaat wat het effect is van deze variabele op een afhankelijke variabele (bijv. vermoeidheid). Een laboratorium is de ideale plek voor een zuiver experiment, maar onderzoekers in de A&G-psychologie kunnen ook een veldexperiment opzetten waarbij de omstandigheden in een organisatie zoveel mogelijk constant worden gehouden. Kerlinger en Lee (2000) verwijzen naar het max/min/con-principe, om aan te geven dat het behalve om controle van externe systematische variantie gaat om maximalisatie van de systematische variantie in de afhankelijke variabele die bestudeerd wordt, en om minimalisatie van de foutenvariantie. Deze drie aspecten worden hier achtereenvolgens behandeld.

#### 10.4.1 Controle van externe variabelen

Wanneer we bijvoorbeeld in experimenteel onderzoek willen nagaan wat het effect is van werkdruk op vermoeidheid, en we denken dat vermoeidheid ook afhankelijk is van sekse, dan kunnen we ervoor kiezen om alleen mannelijke deelnemers te betrekken in het onderzoek. In dat geval elimineren we het effect van sekse. Het probleem hiermee is echter dat de resultaten van het experiment niet meer gegeneraliseerd kunnen worden naar alle werknemers. We kunnen dan niets zeggen over de relatie tussen werkdruk en vermoeidheid bij vrouwen.

Een tweede manier om te controleren voor externe variantie is door randomisatie. Theoretisch gezien is randomisatie de enige manier om te controleren voor alle mogelijk denkbare externe variabelen. Wanneer deelnemers willekeurig worden toegewezen aan experimentele condities (bijv. lage versus hoge werkdruk) kunnen de experimentele groepen als gelijk worden opgevat, mits deze groepen groot genoeg zijn. In dit geval zijn immers deelnemers met specifieke kenmerken (bijv. mensen met een bepaalde stemming of persoonlijkheid) naar verwachting evenredig verdeeld over de condities. Bij veldexperimenten is randomisatie in de regel niet mogelijk, vandaar dat in dergelijke gevallen van quasi-experimenteel onderzoek wordt gesproken. Immers, aan één bepaalde voorwaarde voor experimenteel onderzoek – randomisatie – is niet voldaan.

Een derde manier om te controleren voor een externe variabele, is om deze variabele expliciet op te nemen in het onderzoeksontwerp. In het voorbeeld zou sekse in het design kunnen worden opgenomen naast werkdruk. Hierna kan worden nagegaan of mannen en vrouwen verschillen in vermoeidheid (de afhankelijke variabele) en of zij verschillend reageren op de manipulatie van werkdruk (interactie-effect).

Tot slot is het mogelijk om te kiezen voor matching; hierbij wordt geprobeerd om zoveel mogelijk gebruik te maken van vergelijkbare deelnemers in de verschillende experimentele condities. Dit kan bijvoorbeeld door van tevoren vast te stellen wat de persoonlijkheid van deelnemers is of hoe hun stemming is. Bij de toewijzing aan condities worden daarna deelnemers met vergelijkbare kenmerken toegewezen aan de experimentele condities. Matching is echter in de praktijk vaak lastig wanneer er meer dan twee variabelen in het geding zijn, omdat men dan veelal niet voldoende vergelijkbare deelnemers overhoudt. Bovendien is matching nauwelijks een goed alternatief voor randomisatie omdat het aantal variabelen waarop ‘gematched’ kan worden gering is.

#### 10.4.2 Maximalisatie van de systematische variantie

Bij maximalisatie van de systematische variantie gaat het om het maximaliseren van de experimentele variantie in de afhankelijke variabele, veroorzaakt door de onafhankelijke variabele. In onderzoek naar de relatie tussen werkdruk en vermoeidheid willen we dat de variantie in scores op vermoeidheid zoveel mogelijk kan worden veroorzaakt door de onafhankelijke

variabele werkdruk. Hierbij is het van belang om de afhankelijke variabele vermoeidheid de kans te geven om ‘zichzelf te laten zien’, want wanneer er nauwelijks variatie is in werkdruk, zullen we ook weinig variatie waarnemen in vermoeidheid. Het onderzoek moet dus zo opgezet worden dat de experimentele condities (bijv. lage versus hoge werkdruk) zo verschillend mogelijk zijn. Doel is hier dus om te onderzoeken of een bepaalde onafhankelijke variabele (werkdruk) effect heeft op een bepaalde afhankelijke variabele (vermoeidheid) en niet zozeer om te zien hoe groot dat effect is in alledaagse situaties, in verhouding tot andere verklarende variabelen.

#### 10.4.3 Minimalisatie van foutenvariantie

Foutenvariantie is de variabiliteit in variabelen als gevolg van willekeurige fluctuaties. Variantie in vermoeidheidsscores kan bijvoorbeeld foutief veroorzaakt worden doordat deelnemers proberen om de hypothese te raden, tijdelijk niet opletten, wisselingen laten zien in stemming of tussentijds even rust nemen. Bij het minimaliseren van foutenvariantie gaat het om twee zaken: reductie van meetfouten door gebruik te maken van gecontroleerde condities en vergroting van de betrouwbaarheid van de meetinstrumenten. Hoe meer er sprake is van een ongecontroleerde situatie en onbetrouwbare maten, des te groter is de kans dat de veroorzakers van foutenvariantie hun werk kunnen doen.

Hoewel het lastig is om in veldexperimenteel onderzoek (bijv. bij callcentermedewerkers) te controleren voor foutenvariantie, moet dit natuurlijk toch worden geprobeerd. Foutenvariantie kan worden verlaagd door de deelnemers heel specifiek en duidelijk te instrueren, door hen tijdens het onderzoek te observeren of te controleren (bijv. door video-opnamen te maken of door ‘mee te luisteren’) en door externe factoren zoveel mogelijk uit te schakelen (bijv. aan- of afwezigheid van de leidinggevende tijdens het onderzoek). Bovendien geldt: hoe betrouwbaarder de meetinstrumenten, des te groter de kans dat we daadwerkelijk systematische en gewenste variantie meten.

#### 10.4.4 Sterke en zwakte punten van experimenteel onderzoek

Het sterke punt van laboratoriumexperimenten is de relatief hoge interne validiteit doordat onderzoek wordt verricht onder streng gecontroleerde condities. Wanneer werkdruk wordt gemanipuleerd bij werknemers die willekeurig zijn toegewezen aan experimentele condities (lage werkdruk versus hoge werkdruk), kunnen we er vrij zeker van zijn dat verschillen tussen de groepen op de afhankelijke variabele ‘vermoeidheid’ veroorzaakt zijn door de onafhankelijke variabele ‘werkdruk’. Desondanks kan de interne validiteit net als bij longitudinaal onderzoek ook bij experimenteel onderzoek bedreigd worden door bijvoorbeeld een testeffect, statistische regressie of uitval (voor een overzicht zie tabel 10.1). Door een experimenteel design uit te breiden met een voor- en nameting en een controlegroep, kunnen overigens veel van dit soort problemen worden voorkomen.

Kerlinger en Lee (2000) stellen dat het zwakke punt van laboratoriumonderzoek is ‘probably the lack of strength of independent variables’ (p. 580). In het onderzoek dat Zijlstra en Meijman beschrijven in hoofdstuk 3 werden deelnemers al dan niet moe gemaakt door hen ’s ochtends zes uur achtereen te laten werken aan mentaal belastende examenopgaven (Zijlstra & Meijman, 1989). Voorts moesten zij ’s middags een makkelijke of moeilijke geheugenzoektaak uitvoeren, waarbij ze veel of juist weinig mogelijkheden hadden om hun taaklast (een geheugenzoektaak) te reguleren. Uit het experiment bleek onder andere dat de lage taaklast nauwelijks als inspannender werd beoordeeld dan de hoge taaklast wanneer de deelnemers in een vrij tempo konden werken. Bovendien bleek dat vermoeidheid weinig invloed had op de inspanning die men zich getrooste wanneer men veel regelmatig had, maar juist tot verminderde inspanning leidde wanneer men weinig regelmatig had.

Hoewel deze resultaten opmerkelijk zijn, en een fraaie illustratie vormen van het Demand-Control Model, is het eenvoudig om in te zien dat de effecten van de manipulaties relatief zwak zijn. Vergelijk dit maar eens met de sterke effecten van onafhankelijke variabelen in realistische arbeidssituaties, waarbij chronische vermoeidheid kan leiden tot ziekteverzuim en instroom in de WIA!

De belangrijkste tekortkoming van laboratoriumonderzoek is dus het gebrek aan externe validiteit of generaliseerbaarheid. Dit wordt ook wel aangeduid met de term ecologische validiteit. Zijlstra (1997) pleit voor het vergroten van deze ecologische validiteit door het verrichten van onderzoek in een gesimuleerde maar realistische taakomgeving, in plaats van in een klassiek laboratorium waarbij een sterke reductie van het onderzoeksobject optreedt.

Een voorbeeld van een dergelijke gesimuleerde taakomgeving is het onderzoek van Deelstra e.a. (2003) naar de negatieve effecten van sociale steun, waarbij proefpersonen (secretarissen) een reële taak kregen opgedragen (het plannen van een conferentie) in een als kantoor omgebouwde experimentele ruimte. Tijdens het uitvoeren van de taak kregen de proefpersonen ongewild sociale ‘steun’ van een collega (een handlanger van de proefleider). Zoals verwacht bleek die ongewilde steun een negatief psychologisch en psychofysiologisch effect te hebben.

### 10.5 Objectieve en subjectieve benaderingen

In deze paragraaf wordt aandacht besteed aan methodische overwegingen bij subjectieve dan wel objectieve benaderingen in onderzoek naar oorzaken van werkstress en arbeidsmotivatie. Is werkdruk vooral afhankelijk van de subjectieve waarneming van werknemers of kunnen we het ook objectief vaststellen? Wat is belangrijker, en hoe kunnen we beide conceptueel in een model onderbrengen? Deze vragen zijn uiterst relevant, omdat het meeste onderzoek op het gebied van de A&G-psychologie is gebaseerd op subjectieve metingen van werkkenmerken en werkbeleving. Wanneer er geen relatie

bestaat tussen de werkelijkheid en de representaties ervan in oordelen over de kwaliteit van het arbeidsleven, kan men zich afvragen wat de betekenis is van de onderzoeksresultaten voor de praktijk.

### 10.5.1 Objectieve en subjectieve stressoren

In de literatuur worden verschillende omschrijvingen van subjectief en objectief gebruikt. In dit hoofdstuk volgen we de benadering van Frese en Zapf (1988), die spreken van objectieve stressoren wanneer deze geen voorwerp zijn van cognitieve of emotionele verwerking, en van subjectieve stressoren wanneer dat wel het geval is. Een voorbeeld van een objectieve stressor is de grote hoeveelheid telefoontjes die de teleconsultants van het callcenter dagelijks afhandelen. Het aantal telefoontjes wordt zonder tussenkomst van de teleconsultants (objectief) bijgehouden door de computer. Een voorbeeld van een subjectieve stressor is de ervaren werkdruk door teleconsultants.

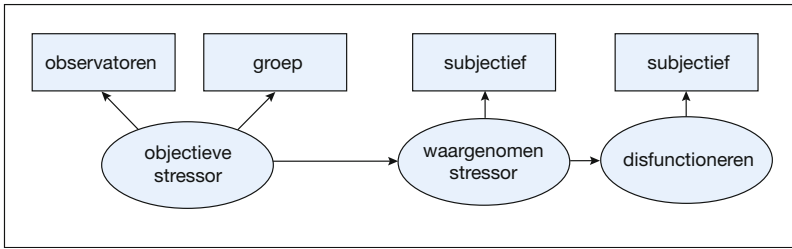
Maar hoe is het nu mogelijk om van objectieve stressoren te spreken terwijl de subjectieve processen die de stressreactie constitueren zijn uitgeschakeld? Frese en Zapf (1988) voeren daartoe een fictieve ‘gemiddelde persoon’ ten tonele. Wanneer deze persoon geacht kan worden om het betreffende werkkenmerk als stressvol te beleven, is er sprake van een objectieve stressor. Zo is het aannemelijk dat het aantal binnenkomende telefoontjes een potentiële stressor voor teleconsultants vormt en de kleur haar van de chef niet. Het is van belang erop te wijzen dat objectieve stressoren potentieel en niet per definitie tot stressreacties leiden.

In navolging van Lazarus en Folkman (1984) gaan we er voorts van uit dat een objectieve stressor eerst moet worden waargenomen voordat deze een psychologisch effect kan hebben op een individuele werknemer. Dit betekent voor bouwvakkers bijvoorbeeld dat langdurige blootstelling aan lawaai wel tot doofheid kan leiden (en daaraan gekoppeld tot disfunctioneren) maar niet noodzakelijk tot stress, doordat ze de langzame achteruitgang van hun gehoor niet als zodanig opmerken. Maar zouden ze zich realiseren dat ze doof worden, dan krijgen ze wel stress.

Werknemers hoeven zich dus niet altijd bewust te zijn van objectieve stressoren. In bepaalde beroepen (bijv. ambulancepersoneel) is het zelfs heel functioneel om het stressvolle karakter van het werk niet doorlopend waar te nemen. Vooral wanneer zij de objectief stressvolle situaties – waarin zij zich dagelijks bevinden – psychologisch interpreteren, kan dit op termijn leiden tot stress. In het algemeen geldt dat stress vooral dan optreedt wanneer men zich realiseert dat de taakeisen de eigen vaardigheden te boven gaan. Dit zogenoemde stressperceptiemodel staat schematisch afgebeeld in figuur 10.2.

### 10.5.2 Werkkenmerken objectief vaststellen

Waarom is het zinvol om in A&G-psychologisch onderzoek te zoeken naar objectieve indicatoren? Volgens Frese en Zapf (1988) zijn er drie typen redenen aan te wijzen: theoretische, methodische en praktische redenen.



*Figuur 10.2*

Stressperceptiemodel.

Bron: Frese en Zapf (1988).

Theoretisch gezien is het een belangrijke vraag hoe perceptie samenhangt met de werkelijkheid. Waar de nadruk in louter subjectieve benaderingen van bijvoorbeeld werkstress en motivatie te veel ligt op de cognities van werknemers, is eenzijdige aandacht voor de objectieve werkomgeving te typeren als mechanistisch, omdat er geen rekening wordt gehouden met de waarneming en interpretatie van mensen ('appraisal'). Het is inhoudelijk belangrijk om na te gaan hoe de objectieve werkomgeving van invloed is op de perceptie ervan. Zo kan bijvoorbeeld objectieve feedback uit de werkomgeving ervoor zorgen dat werknemers meer realistische doelen stellen en hun prestaties beter inschatten. Het gevolg hiervan kan zijn dat men op termijn betere prestaties levert dan wanneer men geen feedback ontvangt.

De methodische reden om gebruik te maken van objectieve indicatoren van de werkomgeving heeft te maken met het probleem dat er triviale correlaties kunnen optreden tussen subjectieve indicatoren van arbeid en gezondheid, bijvoorbeeld als gevolg van gemeenschappelijke methodevariantie of doordat er inhoudelijk overlap bestaat tussen de items van de onafhankelijke (oorzakelijke) en afhankelijke (uitkomst)variabele. Wanneer werknemers de enige bron van informatie zijn en zij worden ondervraagd over zowel de werkdruk als hun vermoeidheid (dus met een 'gemeenschappelijke methode'), is de kans groot dat het statistische verband tussen beide variabelen relatief sterk is.

Volgens Frese en Zapf (1988) kunnen we alleen met objectieve meetinstrumenten vermijden dat we zulke (soms triviale) verbanden gaan duiden. Dit wordt ook wel de 'triviality trap' genoemd. Sommige onderzoekers wijzen er overigens op dat objectieve methoden de relatie tussen twee concepten naar beneden toe kunnen vertekenen, omdat subjectieve en objectieve methoden hun eigen bronnen van (statistisch onafhankelijke) foutenvariantie hebben. Als men één methode gebruikt, zijn de bronnen van foutenvariantie hetzelfde. Het gevolg hiervan kan zijn dat samenhangen naar boven toe vertekend zijn. (Voor een uitgebreide bespreking zie Podsakoff e.a., 2003.)

De praktische reden om onderzoek te doen naar objectieve stressoren in de werkomgeving is dat onderzoekers en organisaties vaak op grond van onderzoek de stressoren in het werk willen reduceren en de hulpbronnen

willen vergroten. Dit kan in principe op twee manieren worden gerealiseerd: door de werkcontext aan te passen of door het individu aan te passen. Het heeft uiteraard alleen zin om de kenmerken van de werkorganisatie te veranderen wanneer inderdaad blijkt dat objectieve werkstressoren leiden tot gezondheidsklachten en ziekteverzuim.

Als een stressor zoals werkdruk slechts bestaat in de ogen van de werknemer ('in the eye of the beholder') en afhankelijk is van de individuele waarneming en interpretatie van de situatie, is het niet noodzakelijk om de werkomgeving te veranderen. In dat geval kan het ook zinvol zijn om de percepties van het individu te veranderen. Frese en Zapf (1988) wijzen er dan ook op dat de vraag of objectieve stressoren bestaan belangrijke strategische implicaties heeft. Immers, een functieherontwerpbenadering (zie hfst. 9 en 11) zou weggegooid geld zijn wanneer werkomstandigheden louter subjectief zijn.

### 10.5.3 Objectieve en subjectieve stressoren meten

In A&G-psychologisch onderzoek wordt er doorgaans van uitgegaan dat subjectieve methoden verwijzen naar vragenlijsten die worden ingevuld door werknemers, terwijl objectieve methoden verwijzen naar beoordelingen door deskundigen, documentenanalyses en fysiologische metingen (zie box 10.2). Strikt gesproken is deze veelgebruikte indeling niet juist. Frese en Zapf (1988) geven het voorbeeld van een respondent die door middel van een vragenlijst aangeeft dat hij 25 jaar is. In dit geval zijn onderzoekers vaak geneigd om aan te nemen dat de gegevens objectief zijn. Wanneer het daarentegen gaat over attitudes ten aanzien van het werk, nemen we aan dat er veel meer cognitieve verwerking heeft plaatsgevonden om tot een oordeel te komen. Kortom, elke rapportage via een vragenlijst kan worden geschaald op een dimensie die loopt van 'weinig afhankelijk van cognitieve en emotionele verwerking' tot 'zeer afhankelijk van cognitieve en emotionele verwerking' (zie ook hfst. 5).

#### Box 10.2 Voorbeelden van objectieve werkkenmerken

##### *Documenten*

Gorgievski-Duijvesteyn (2002) deed onderzoek bij ongeveer vijfhonderd boeren en tuinders naar de relatie tussen financiële problemen en gezondheid. Om financiële problemen te kunnen vaststellen, werd de hulp ingeschakeld van het Landbouw Economisch Instituut. Dit instituut stelt jaarlijks vast hoe kredietwaardig en modern boerenbedrijven zijn. Deze gegevens over de solvabiliteit, liquiditeit en moderniteit werden samengevoegd in een index voor objectieve financiële problemen. De objectieve index bleek rond de 0,30 te correleren met de subjectieve perceptie van de boeren van hun eigen financiële situatie, en indirect voorspellend te zijn voor hun mentale gezondheid gedurende de twee jaar volgend op de vaststelling van de financiële cijfers.



*Oordelen van deskundigen*

Meijman en Van Ouwerkerk (1999) vroegen zich af of anderen zien wat wij doen op ons werk en deden onderzoek bij 54 werknemers in 27 verschillende functies. Zij verzamelden bij deze werknemers subjectieve oordelen over verschillende kenmerken van de werksituatie (bijv. werkdruk, emotionele taakeisen, gebrek aan sociale steun van collega's en gebrek aan autonomie). Bovendien werden twee beoordelaars getraind in het observeren van deze werkkenmerken. De observatoren liepen tweemaal een halve dag rond op de werkplek, en hielden gesprekken met zowel de werknemers als hun direct-leidinggevenden. De correlaties tussen de objectieve en subjectieve oordelen waren het hoogst voor werkstressoren zoals werkdruk en emotionele taakeisen (rond de 0,50) en duidelijk lager voor hulpbronnen zoals sociale steun en coaching (rond de 0,25). Deze laatste aspecten hebben een meer sociaal karakter en zijn door een externe beoordelaar kennelijk moeilijker te observeren.

De problemen bij subjectieve rapportages van werkstressoren en werkbeleving worden in handboeken over methoden en technieken van onderzoek uitgebreid besproken. Frese en Zapf (1988) geven verschillende voorbeelden.

- Meetfouten die in de klassieke testtheorie worden besproken, bijvoorbeeld de geneigdheid om extreme antwoorden te geven of juist te vermijden en het schaal midden als antwoord te kiezen. Of het halo-effect: de neiging om alle specifieke oordelen over een onderwerp (bijv. de werksituatie) te laten beïnvloeden door de algemene indruk ervan.
- Inhoudelijke overlap tussen onafhankelijke en afhankelijke variabelen.
- Problemen met een derde variabele die zowel de onafhankelijke als de afhankelijke variabele beïnvloedt (bijv. een persoonlijkheidstrekk of de neiging om te klagen ofwel NA; zie par. 10.6).
- De invloed van het persoonlijk welbevinden op de beoordeling van stressoren.
- ‘Demand characteristics’: kenmerken van het onderzoek die de deelnemers aanmoedigen om aan onderzoekers te rapporteren wat ze worden verondersteld te willen.

De oordelen van externe beoordelaars zijn uiteraard net als die van werknemers zelf afhankelijk van de cognitieve en emotionele verwerking van informatie. Ook beoordelaars kunnen worden beïnvloed door hun eigen welbevinden of persoonlijkheid. In het geval van beoordelaars worden er in de beoordeling echter fouten gemaakt, die onafhankelijk zijn van de oordelen van de werknemers zelf over hun eigen (dis)functioneren. Frese en Zapf (1988) laten zien dat het gevolg hiervan is dat de correlatie tussen werkkenmerken en het (dis)functioneren systematisch wordt onderschat.

Tot slot zijn er verschillende specifieke problemen met het gebruik van oordelen van observatoren (Frese & Zapf, 1988). Allereerst is de tijd om te observeren om praktische redenen vaak beperkt. De werkzaamheden die

men kan observeren in een kort tijdsbestek vormen daardoor slechts een beperkte afspiegeling van de realiteit. Als gevolg hiervan is de kans groot dat observatoren bijvoorbeeld pieken in werkdruk niet waarnemen, terwijl vooral deze wellicht verantwoordelijk zijn voor werkstress. Ten tweede moeten observatoren soms een inschatting geven van stressoren die sterk verweven zijn met mentale processen (bijv. hoge concentratie). Zulke mentale processen kunnen niet worden waargenomen en het is dus waarschijnlijk dat er fouten in de beoordelingen sluipen.

Ten derde werken mensen anders wanneer ze weten dat zij geobserveerd worden. Ze doen het iets rustiger aan, nemen de tijd om dingen uit te leggen aan de observator, en voeren taken uit waar ze de voorkeur aan geven. Men houdt zich vaak ook beter aan veiligheidsvoorschriften bij observatie. Tot slot kent elk onderzoek z'n praktische beperkingen. Een beperking van observatieonderzoek is dat beoordelaars in bepaalde gevallen geen toegang hebben tot organisaties waar de arbeidsomstandigheden suboptimaal zijn. Zelfs in organisaties die vrijwillig meewerken kan er een neiging bestaan om onderzoekers alleen toegang te geven tot werkplekken die het visitekaartje zijn van de organisatie.

Kort samengevat blijkt dus dat objectieve metingen in het algemeen resulteren in een onderschatting van effecten en dat subjectieve oordelen doorgaans leiden tot een overschatting. Het is uiteraard van belang om in onderzoek rekening te houden met dit gegeven. Voorts is het nuttig om in onderzoek zowel subjectieve als objectieve indicatoren op te nemen. Tezamen geven zij een gedetailleerder en genuanceerder inzicht in relaties dan alleen subjectieve of alleen objectieve indicatoren. Bovendien kan men met behulp van meer bronnen van informatie triangulatie toepassen op de gegevens. Dat wil zeggen dat men kan nagaan wat het verband is tussen theoretisch vergelijkbare concepten, die op verschillende manieren zijn ge-operationaliseerd.

## 10.6 Derde variabelen

In het veld van de A&G-psychologie komt het vaak voor dat een derde variabele Z invloed heeft op de relatie tussen twee variabelen X en Y. We bespreken hier drie van deze invloeden, die bekend staan als het 'derde-variabeleprobleem', mediatoreffecten en moderatoreffecten.

### 10.6.1 Derde-variabeleprobleem: negatieve affectiviteit

Om te kunnen spreken van een causale relatie tussen twee variabelen, mag de samenhang tussen deze variabelen niet te wijten zijn aan de invloed van een 'derde' variabele (zie figuur 10.3a). Mensen die een hoge werkdruk ervaren, maken bijvoorbeeld vaak fouten op het werk. Dat betekent echter niet dat een hoge werkdruk altijd leidt tot het maken van fouten. Hoewel het maken van fouten heel goed kan leiden tot een hogere werkdruk (de causale relatie ligt dan omgekeerd), is het denkbaar dat vooral stressgevoelige werknemers

een hoge werkdruk zullen ervaren én meer fouten zullen maken dan anderen. De samenhang tussen beide verschijnselen kan dus in ieder geval ten dele het gevolg zijn van een gemeenschappelijke derde variabele, namelijk individuele stressgevoeligheid.

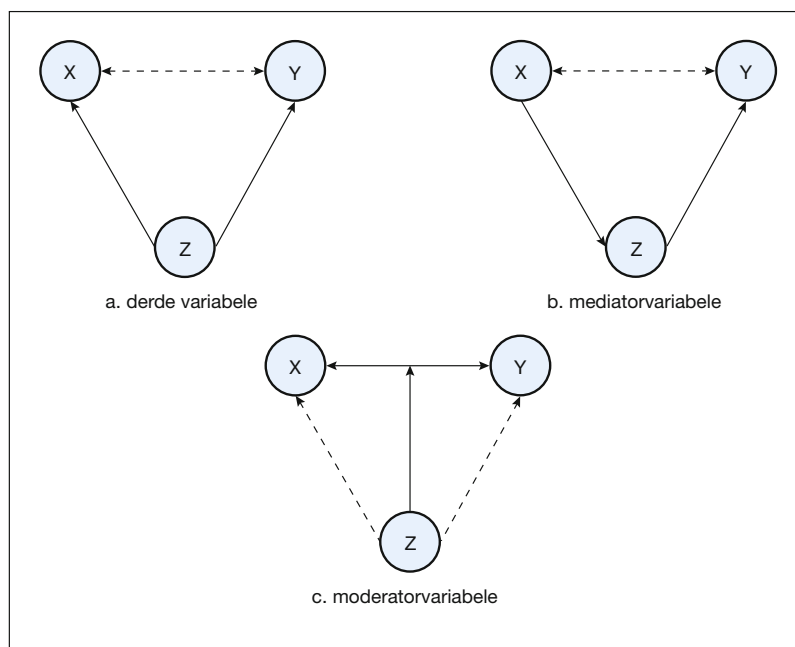
Dergelijke alternatieve verklaringen voor een samenhang tussen twee variabelen moeten natuurlijk zoveel mogelijk worden uitgesloten. Dat vereist allereerst dat onderzoekers weten welke variabelen potentieel een rol als derde variabele kunnen vervullen. Het in de A&G-psychologie wellicht bekendste voorbeeld van zo'n derde variabele is negatieve affectiviteit (NA; Watson & Pennebaker, 1989; zie ook hfst. 5), een stabiele persoonlijkheidstrekk. Mensen met een hoge score op NA hebben de neiging om alles wat er in hun omgeving en met henzelf gebeurt in een negatief licht te zien. Die neiging zal zich ook doen gelden bij het invullen van vragenlijsten of het geven van antwoorden tijdens een interview. Callcentermedewerkers met een hoge score op NA zullen bijvoorbeeld zowel een hogere werkdruk als meer vermoeidheid rapporteren. Het verband tussen deze concepten zal dan op zijn minst ten dele worden veroorzaakt door negatieve affectiviteit, en interpretatie van dit verband in causale termen is dan niet zonder risico.

In A&G-psychologisch onderzoek naar deze kwestie is aangetoond dat NA inderdaad de sterkte van relaties tussen concepten kan beïnvloeden (Spector, 2006). Dat is echter niet hetzelfde als te zeggen dat NA deze relaties vertekent. NA zou volgens Spector e.a. een inhoudelijk goed interpreteerbare rol kunnen spelen in het stressproces, bijvoorbeeld als mediator tussen werkdruk en gezondheidsklachten (zie verder). Het lijkt bijvoorbeeld aannemelijk dat mensen die veel werkdruk ervaren, als gevolg daarvan de neiging zullen hebben het werk en alles wat daarmee samenhangt negatief te beoordelen. In zo'n optiek is NA niet zozeer een verstorende variabele maar eerder een uitkomstmaat, op dezelfde manier als uitputting en lichamelijke klachten dat zijn.

Op dit moment is nog onduidelijk welke betekenis men in de A&G-psychologie moet toekennen aan de effecten van NA. Een voorzichtige aanbeveling is om in onderzoek altijd een lijst voor negatieve affectiviteit op te nemen, zoals de Positief en Negatief Affect Schaal (PANAS; Peeters, Ponds & Boon-Vermeeren, 1999). Op die manier kan altijd empirisch worden nagegaan hoe groot de effecten van NA eigenlijk zijn en of de sterkte van relaties inderdaad wezenlijk verandert als er voor NA wordt gecontroleerd. Is dat laatste niet het geval, dan kan NA als verklaring voor in de studie gevonden verbanden worden weggestreept. Zijn er wel duidelijke verschillen zichtbaar, dan is het zinvol om verder na te denken over de rol die NA inhoudelijk zou kunnen spelen.

### 10.6.2 Mediatoreffecten

Er is sprake van mediatie wanneer een variabele X van invloed is op Y, via variabele Z (zie figuur 10.3b). Stel dat een onderzoeker de theorie heeft opgesteld dat werkdruk vooral van invloed is op ziekteverzuim, via de spanningen die dat oproept. Aangenomen wordt dat werkdruk resulteert in



Figuur 10.3

Verschillende mogelijkheden van beïnvloeding van de relatie tussen variabelen X en Y door variabele Z.

spanningen, en spanningen leiden op hun beurt weer tot ziekteverzuim. Strikt genomen is het uitgangspunt dat de relatie tussen werkdruk en verzuim tot stand komt door het optreden van spanningen. Spanningen spelen zo'n mediërende of interveniërende rol wanneer aan de volgende voorwaarden is voldaan:

- de werkdruk hangt significant samen met zowel spanningen als met ziekteverzuim;
- de spanningen vertonen een significante samenhang met ziekteverzuim;
- de correlatie tussen werkdruk en verzuim verdwijnt (of vermindert significant) na controle voor de correlatie van beide variabelen met spanningen.

Anders gezegd: de invloed van de mediërende variabele 'spanningen' kan worden vastgesteld door het berekenen van de partiële correlatie tussen werkdruk en verzuim, na het verwijderen van de correlatie die is toe te schrijven aan hun wederzijdse samenhang met spanningen (zie Baron & Kenny, 1986).

### 10.6.3 Moderatoreffecten

Er is sprake van een moderatoreffect wanneer de relatie tussen variabele X en Y varieert voor verschillende groepen (werknemers die variëren op aspect Z;

zie figuur 10.3c). Volgens het Demand-Control Model van Karasek (zie hfst. 2) zijn regelmogelijkheden (Z) de moderator van het ongewenste effect van werkdruk (X) op ziekteverzuim (Y). Voor werknemers met weinig autonomie gaat een toename in werkdruk gepaard met een toename in verzuim. Voor werknemers met veel regelmogelijkheden is er geen (of een veel minder sterke) samenhang tussen werkdruk en verzuim.

Met andere woorden, het DC Model veronderstelt dat regelmogelijkheden een buffer kunnen zijn tegen het effect van werkdruk op stressreacties. In een regressievergelijking ter voorspelling van verzuim kan dit worden vastgesteld door een significante bijdrage van de interactieterm XZ (werkdruk maal regelmogelijkheden). Om inzicht te verkrijgen in de richting van het interactie-effect kan men de correlatie tussen werkdruk en verzuim uitrekenen voor de groep met weinig regelmogelijkheden en de groep met veel regelmogelijkheden. Het model van Karasek wordt bevestigd wanneer de correlatie tussen werkdruk en verzuim significant en positief is voor de eerste groep en wanneer deze relatie zwakker of zelfs afwezig is voor de tweede groep.

### 10.7 Meta-analyse

Doel van een meta-analyse is de resultaten van verschillende studies naar een bepaald verschijnsel met elkaar in verband te brengen en te integreren. Daarbij is de fundamentele veronderstelling dat de resultaten van afzonderlijke studies in meer of mindere mate vertekend zullen zijn door allerlei toevalsfactoren. Wanneer al deze studies met elkaar gecombineerd worden, zullen deze toevalsfactoren echter tegen elkaar wegvallen. Daardoor kan een zuiverder beeld worden verkregen van de relatie tussen twee concepten dan mogelijk zou zijn in één studie. Tijdens een meta-analyse worden achtereenvolgens vier stappen gezet.

*Omschrijven van het te onderzoeken verschijnsel of verband.* In deze stap wordt het te onderzoeken verschijnsel zo precies mogelijk gedefinieerd. Dat kan heel ver gaan: twee instrumenten om personeelsverloop te meten kunnen bijvoorbeeld inhoudelijk sterk verschillen (bijv. verloopintentie versus feitelijk, ‘objectief’ gemeten verloop), waardoor de resultaten die verkregen zijn met het ene instrument anders kunnen zijn dan de resultaten van het andere instrument. Het kan dan zinvol zijn de meta-analyse te beperken tot een bepaald type uitkomstmaat (bijv. het feitelijke verloop).

*Verzamelen van studies waarin het betreffende verschijnsel of verband is onderzocht.* De meest voor de hand liggende manier om dergelijke studies op te sporen is door gebruik te maken van geautomatiseerde zoeksystemen als PsycINFO of MEDLINE. Via dergelijke systemen is de wetenschappelijke literatuur gemakkelijk te ontsluiten. Andersoortige literatuur (bijv. dissertaties, interne rapporten) is vaak moeilijker op te sporen, maar ook veelal de moeite waard.

*Studies selecteren.* Niet alle in stap 2 opgespoorde studies zullen van vergelijkbare kwaliteit zijn. In sommige studies worden afwijkende of zelfs slechte instrumenten gebruikt, in andere studies is sprake van een zeer kleine steekproef of een zeer hoge non-respons, enzovoort. Dergelijke studies zijn minder of zelfs niet bruikbaar en kunnen om die reden dan ook uitgesloten worden.

*Resultaten van de geselecteerde studies samenvatten.* De studies worden beschreven in termen van onderzoeksgroep, type instrumenten, resultaten en factoren die de uitkomsten van de studie zouden kunnen beïnvloeden. Vervolgens worden de ‘stemmen’ geteld. Hoeveel studies vinden een positief verband? Hoeveel een negatief? Hoeveel studies vinden geen verband? In de meta-analyse wordt daarnaast allerlei numerieke en kwantitatieve informatie betreffende de in de studie gevonden verbanden opgenomen. Daarbij kan het gaan om grootheden als *t*- en *F*-waarden (die de statistische significantie van bepaalde verbanden weergeven) of om grootheden die betrekking hebben op de sterkte van bepaalde verbanden (bijv. de grootte van correlaties). Beide typen grootheden kunnen via allerlei rekenregels naar elkaar worden omgerekend (zie bijv. Hunter & Schmidt, 1990). Op die manier kunnen de in verschillende studies gevonden verbanden tussen bepaalde concepten onder één noemer worden gebracht, zodat vergelijking gemakkelijk is. Als de effecten van verschillende studies in dezelfde grootheid zijn uitgedrukt, kunnen de resultaten van deze studies ook worden uitgedrukt in één getal, dat indicatief is voor de sterkte van het verband tussen de onderzochte concepten. In een door Irvine en Evans (1995) uitgevoerde meta-analyse naar verloop en satisfactie onder verpleegkundigen bleek bijvoorbeeld dat er zowel studies waren die een negatief verband tussen satisfactie en verloop rapporteerden, als studies waarin geen of een positief verband werd gevonden. Nadat al het empirische bewijs op een rijtje was gezet, bleek dat er per saldo slechts een klein, negatief verband tussen satisfactie en verloop aanwezig was.

De meta-analyse is een bruikbare methode om de resultaten van eerder onderzoek samen te vatten. Er kan een redelijk precieze schatting gemaakt worden van de sterkte van een bepaald verband. Dergelijke statistische exercities zijn echter vooral zinvol als er veel studies naar een bepaald verschijnsel zijn verricht, en dat is in een relatief jonge tak van wetenschap als de A&G-psychologie niet altijd het geval. Voor een uitgebreide bespreking van meta-analyses verwijzen we naar het artikel van Born en Mol (2006).

## 10.8 Multiniveau-onderzoek

Veel onderzoek binnen de A&G-psychologie heeft te maken met analyse-eenheden op verschillende – hiërarchische – niveaus. Zo kennen we een hiërarchische ordening van organisaties, units, afdelingen, teams en indivi-

---

\* Om deze reden is het wenselijk om in elk artikel een correlatiematrix op te nemen.

duale werknemers. Stel dat in een studie zowel informatie over individuen (bijv. individuele werkprestatie) als over teams (bijv. groepsprestatie) en afdelingen (bijv. type afdeling en afdelingsgrootte) is verzameld. Onderzoekers zouden in dat geval bijvoorbeeld graag het effect van type afdeling, afdelingsgrootte en groepsprestatie op de individuele prestatie willen bestuderen. Een dergelijk probleem toont de behoefte aan analysetechnieken die informatie op diverse hiërarchische niveaus aankunnen. Multiniveau-analyse is zo'n techniek die optimaal gebruik maakt van een dergelijke datastructuur, te weten individuen binnen teams en teams binnen afdelingen.

Algemeen beschouwd is multiniveau-analyse een soort multi-pele-regressieanalyse die rekening houdt met meer aggregatieniveaus in de onderzoeksgegevens. Verder biedt de techniek de mogelijkheid om de variantie op individueel niveau (bijv. arbeidsprestatie) op te splitsen in meer – hiërarchische – delen: een deel dat toegeschreven wordt aan verschillen in teams (in dit geval groepsprestatie) en een deel dat toegeschreven wordt aan verschillen in afdelingen (type en grootte).

De vraag resteert waarom we multiniveau-analyse als techniek nodig hebben. Conventionele statistische analysetechnieken zoals regressieanalyse en variantieanalyse gaan uit van de veronderstelling dat alle observaties in een studie onafhankelijk van elkaar zijn. Maar is dat ook zo bij multiniveau-onderzoek? Men kan zich voorstellen dat werknemers binnen dezelfde organisatie meer op elkaar lijken dan op werknemers van een andere organisatie, bijvoorbeeld door selectieprocessen en cultuurverschillen. In dat geval zijn de waarnemingen niet meer onafhankelijk van elkaar en zijn de statistische toetsingen met conventionele technieken mogelijk niet betrouwbaar. Een ander probleem is dat in zeer kleine groepen de groepsgemiddelden grote standaardafwijkingen hebben. Conventionele technieken zijn niet in staat om dergelijke systematische variantie binnen de groepen te onderscheiden van steekproefruis. Binnen multiniveau-analyse kan men echter wel een onderscheid maken in steekproefruis tussen groepen en systematische variantie binnen groepen.

Een groot voordeel van multiniveau-analyse is ten slotte haar nut voor longitudinaal onderzoek (bijv. panelsurveys en dagboekstudies). Multiniveau-analyse is namelijk in staat zeer flexibel om te gaan met panelstudies waarin de onderzoeksgegevens van sommige respondenten niet compleet zijn, of met longitudinale onderzoeksbestanden waar sommige (of zelfs alle) respondenten zijn gemeten op verschillende momenten in de tijd (bijv. met dagboeken). Vanzelfsprekend kent multiniveau-analyse nog meer voordelen, maar het voert in de context van dit hoofdstuk te ver om deze allemaal uit te leggen. Voor een uitvoerige bespreking van multiniveau-analyse wordt verwezen naar gedegen handboeken zoals van Snijders en Bosker (1999) en van Hox (2002).

**Aanbevolen literatuur**

- Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: design and analysis for field settings*. Boston: Houghton Mifflin.
- Frese, M., & Zapf, D. (1988). Methodological issues in the study of work stress: objective vs. Subjective measurement of work stress and the question of longitudinal studies. In C.L. Cooper & R. Payne (Eds.), *Causes, coping and consequences of stress at work* (pp. 375-411). Chichester: Wiley.
- Kerlinger, F.N., & Lee, H.B. (2000). *Foundations of behavioral research* (4e dr.). Orlando, FL: Harcourt.
- Taris, T.W. (2000). *A primer in longitudinal data analysis*. Londen: Sage.