

Samengevoegde datasets: een analytisch instrument voor evidence-based management

'Niet alles wat meetbaar is, is ook belangrijk – en niet alles wat belangrijk is, wordt met zekerheid gemeten', heeft Albert Einstein ooit gezegd. Een inleiding in de instrumenten, methoden, valkuilen, voordelen en veel voorkomende resultaten van evidence-based managementprogramma's met behulp van samengevoegde datasets.

Effectief evidence-based management vereist een analyse van gegevens uit allerlei bronnen² en zorgvuldig ontworpen pretest/posttest-vergelijkingen.³ In onze ervaring tillen echter maar weinig ondernemingen dit proces naar een hoger plan door datasets samen te voegen en daar vervolgens rigoureuze pretest-posttestvergelijkingen en zinnige statistische analyses op los te laten. Als gegevens uit verschillende losse bronnen in een onderneming met elkaar worden samengevoegd, kunnen onderzoekers beslissingen nemen op basis van harde feiten en met een tot nog toe ongekennde precisie, snelheid en reikwijdte proefprojecten uitvoeren. Evidence-based management wordt bij uitstek waardevol als onderzoekers grote bestanden van samengevoegde gegevens opbouwen die gaandeweg steeds meer aan elkaar worden gekoppeld en waarin een tijdreeks aan metingen voor het prestatievermogen uit het verleden, het heden en daarna is opgenomen. Dit artikel biedt de nodige uitleg en achtergrond voor onderzoekers die dergelijke samengevoegde datasets zonder hulp van buitenaf willen opbouwen. Onderzoekers op het gebied van de psychologie,⁴ de geneeskunde en de zorg,⁵ het onderwijs,⁶ de volksgezondheid,⁷ computerwetenschap,⁸ het bedrijfsleven,⁹ en allerlei interdisciplinaire terreinen gebruiken en bepleiten al tientallen jaren aspecten van een evidence-based besluitvorming – vaak, maar niet altijd, onder verwijzing naar de gerespecteerde tradities waar die benadering uit voortkomt: de quasi-experimentele analyse¹⁰ in de gedragswetenschappen en evidence-based geneeskunde.¹¹ Eerst een kort woord over de

theorie die ten grondslag ligt aan evidence-based management en de oorsprong ervan, dit om onze verbreding van het onderzoek in de juiste context te plaatsen.

Theorie en oorsprong in het kort

De theoretische grondslagen van onze benadering vormen een interessant verhaal dat zelden wordt erkend. Zowel in naam als methode komt evidence-based management voort uit 'evidence-based medicine' – een terrein dat meestal wordt toegerekend aan een los verband van artsopleiders. Het eerste rapport van deze groep wordt vaak gezien als de start voor 'evidence-based medicine'.¹² Slechts weinigen beseffen echter dat 'evidence-based medicine' werd ontwikkeld om de artsopleiding te verbeteren en dat de veronderstellingen waar de methode op gebaseerd is, voortkomen uit een theorie over kennisontwikkeling bij volwassenen geformuleerd door Neame en Powis.¹³ Veel onderzoek naar evidence-based management¹⁴ heeft een uitgesproken a-theoretisch karakter (evenals het gerelateerde onderzoek in quasi-experimentele analyse, analytics¹⁵ en business intelligence¹⁶), maar soms is het toch nuttig om voor ogen te houden dat evidence-based management, evenals de medische voorloper ervan, gebaseerd is op theorievorming over hoe volwassenen leren.

Samengevoegde bestanden: het benodigde instrument

Het cruciale instrument bij evidence-based management is een groot samengesteld bestand aan

Palmer Morrel-Samuels doceert enquêteontwerp en onderzoeksmethoden aan de universiteit van Michigan in Ann Arbor. Hij is tevens ceo van Employee Motivation and Performance Assessment (EMPA) en president van de Workplace Research Foundation. Ed Francis is directeur E-Learning bij Root Learning Inc. in Sylvania (Ohio). Steve Shucard was manager in de IT- en luchtvaartindustrie en is nu verbonden aan de Workplace Research Foundation.¹

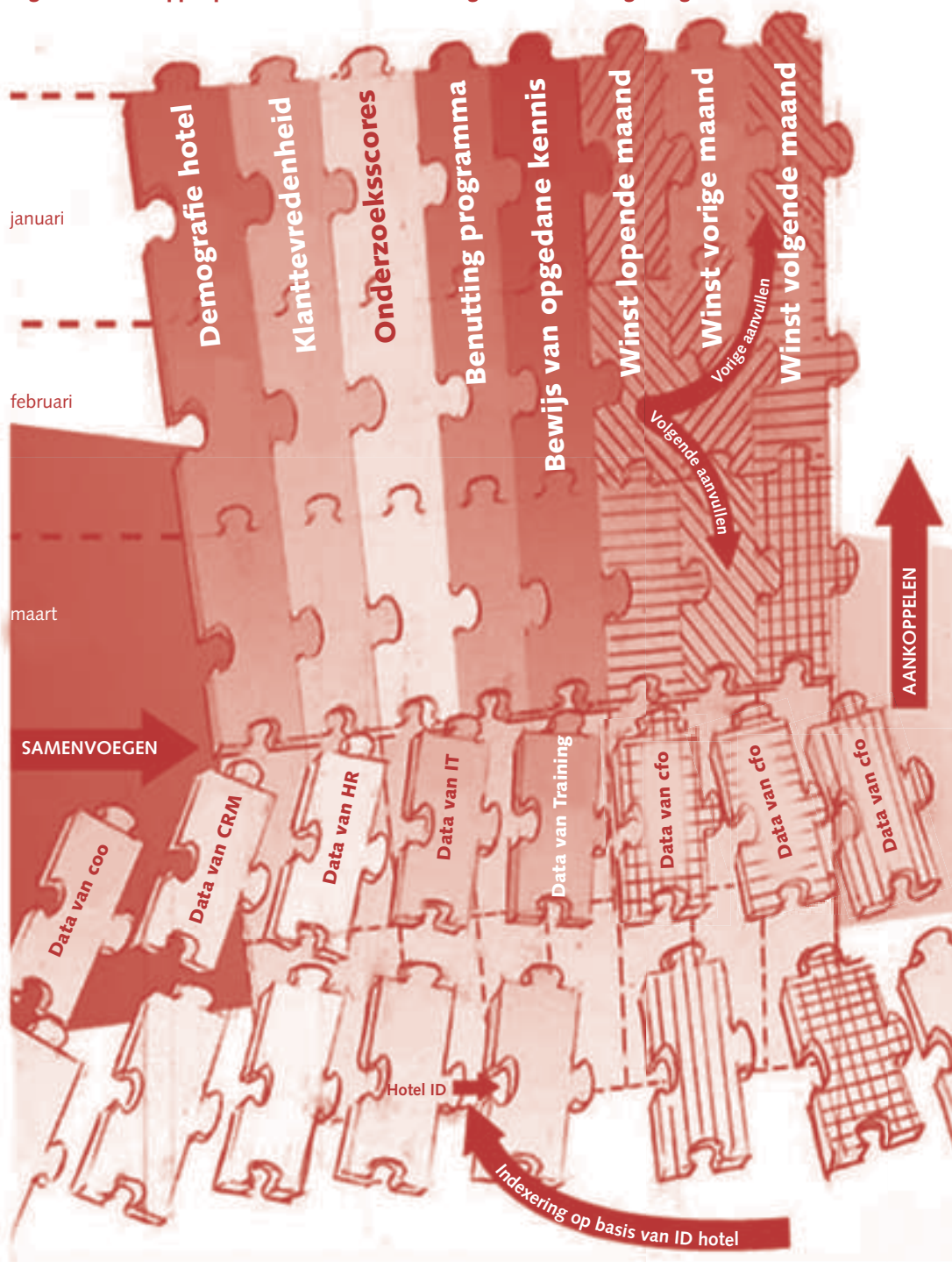
1. De auteurs zijn collega's en cliënten erkentelijk voor hun onontbeerlijke hulp bij het onderhavige onderzoek. Palmer Morrel-Samuels is te bereiken via palmer@umich.edu.

2. J. Pfeffer en R.I. Sutton, 'Management Half-Truths and Nonsense: How to Practice Evidence-Based Management', *California Management Review*, 48/3, Lente 2006, p. 77-100.

3. J. Pfeffer en R.I. Sutton, 'Evidence Based-Management', *Harvard Business Review*, 84/1, januari 2006, p. 62-74.

4. C.F. Michaels, R. Arzamarski, R.W. Isenhowe en D.M. Jacobs, 'Direct Learning in Dynamic Touch', *Journal of Experimental Psychology. Human Perception and Performance*, 34/4, 2008, p. 944-957.
5. S.C. Wangberg, 'An Internet-Based Diabetes Self-Care Intervention Tailored to Self-Efficacy', *Health Education Research*, 23/1, 2008, p. 170-179; L. Bergthold, S. Olson Koebler en S. Singer, 'In Loco Parentis? The Purchaser Role in Managed Care', *California Management Review*, 43/1, Herfst 2000, p. 34-52.
6. P.D. Mautone en R.E. Mayer, 'Cognitive Aids for Guiding Graph Comprehension', *Journal of Educational Psychology*, 99/3, 2007, p. 640-652.
7. Y. Yousey, J. Leake, M. Wdowik en J.K. Janken, 'Education in a Homeless Shelter to Improve the Nutrition of Young Children', *Public Health Nursing*, 24/3, mei/juni 2007, p. 249-255.
8. R. Henderson, F. Deane, K. Barrelle en D. Mahar, 'Computer Anxiety: Correlates, Norms and Problem Definition in Health Care and Banking Employees using the Computer Attitude Scale', *Interacting with Computers*, 7/2, juni 1995, p. 181-193.

Figuur 1. Vierstappenproces voor de ontwikkeling van een samengevoegde dataset



gegevens waarin allerlei 'harde' graadmeters voor het prestatievermogen en 'zachte' enquêtegegevens voor het meten van de bedrijfscultuur aan elkaar worden gekoppeld. Een bekende en rudimentaire toepassing van een dergelijk bestand is om overall in de organisatie het prestatieniveau te meten en veranderingen te registreren. Meer geavanceerde en minder vaak voorkomende toepassingen zijn effectmeting van programma's,¹⁷ verbanden op-

sporen en kwantificeren,¹⁸ positieve afwijkingen benutten,¹⁹ nieuwe trends signaleren die schuil gaan achter de 'achtergrondruis' van niet ter zake doende factoren,²⁰ beoordelen van het effect van bepaalde stijlen van leiderschap,²¹ meten in hoeverre communicatie de winst beïnvloedt,²² en het rendement op de investering (ROI) berekenen van complexe interventies waarbij allerlei variabelen tegelijkertijd invloed uitoefenen.²³ Het primaire

proces bestaat uit rigoureuze, systematische pre-test-posttestvergelijkingen die veel lezers zullen associëren met medisch onderzoek, de volksgezondheid of gedragswetenschap.

De eerste stap op weg naar een samengevoegde dataset ('merged dataset') is om alle belangrijke gegevensbestanden ('databases') voor het meten van het prestatieniveau, de middelen, de winst en de kosten van de onderneming te inventariseren en aan elkaar te koppelen ('merge' of 'join' in automatiseringstaal),²⁴ ongeacht hun reikwijdte, locatie en aandachtsgebied. Dit betekent dat de betreffende gegevens vanuit de hele onderneming in één opslagplaats bijeen worden gebracht, waarbij elke databank wordt geaggregeerd (gemiddeld) door middel van een gemeenschappelijke indexvariabele gebaseerd op een algemene eenheid van analyse (bijvoorbeeld het ID-cijfer van de organisatie) en een kruisverwijzing op tijd (uur, dag, week, maand of jaar), zodat alle informatie kan worden geïndexeerd op datum en bedrijfsonderdeel in de onderneming.

Op gezette tijden (bijvoorbeeld iedere week) worden er onderaan het bestand aanvullende regels met gegevens aangekoppeld. Het gegevensbestand groeit ook doordat er nieuwe kolommen aan worden toegevoegd voor aanvullingen achteraf voor het prestatievermogen in de voorbije en volgende maand. Met andere woorden, er worden vier bewerkingen uitgevoerd (samenvoegen, indexeren, aankoppelen en aanvullen) om het integrale gegevensbestand samen te stellen. Aanvullen is overigens een tweeledig proces: dit betreft zowel actuele gegevens over het verleden (zodat het prestatieniveau van de lopende maand kan worden vergeleken met dat van de vorige) als voor de toekomst (zodat het prestatieniveau van de lopende maand kan worden vergeleken met dat van de volgende). Deze vier bewerkingen zijn niet ingewikkeld, zoals figuur 1 illustreert.

Wie niet goed bekend is met kwantitatieve methoden, denkt misschien dat een onderneming nu ook al gegevens samenvoegt in de winst- en verliesrekening. Op het eerste gezicht lijkt dat ook het geval te zijn. Het dashboard of de scorecard van een gewone samenvatting van de winst- en verliesrekening is echter louter beschrijvend van aard, genereert geen verenigd gegevensbestand, en kan niet worden gebruikt om oorzakelijke verbanden te meten. In tegenstelling tot een gewone samenvatting van een winst- en verliesrekening biedt een analyse van een samengevoegde dataset bijzonder informatieve diagnostiek en beschrijvende informatie

afkomstig uit alle onderdelen van de onderneming. Met behulp van een correct gebruikte samengevoegde dataset kan men initiële oorzaken opsporen, gevolgen meten, de effectiviteit van initiatieven van de onderneming beoordelen, interventies aanbevelen en prognoses maken voor het prestatieniveau. De principes die ten grondslag liggen aan dit type analyse zijn niet nieuw; vaak gaat het erom, bijvoorbeeld, de juiste co-variabelen voor een multivariate statistische analyse te vinden en toe te passen.²⁵ Wel nieuw, echter, zijn het nut, de precisie, het gemak en de reikwijdte – zeker in het bedrijfsleven, waar de analytische methodologie zich minder ver heeft ontwikkeld dan in de gedragswetenschappen.

Samengevoegde datasets: twee beknopte voorbeelden

Samengevoegde datasets kunnen helpen bij de besluitvorming in zeer uiteenlopende situaties, zelfs wanneer de omstandigheden allesbehalve ideaal zijn. Zo is bijvoorbeeld recentelijk met behulp van een samengevoegde dataset een contractueel conflict opgelost tussen Tower, één van de grootste fabrikanten van chassis in de VS, en Lamb, een fabrikant van lasrobots. Tower had bij Lamb een schadeclaim ingediend voor 36 miljoen dollar omdat veelvuldige productieonderbrekingen te wijten zouden zijn aan gebreken in de lasrobots van Lamb. Lamb wist de aanklacht met succes te pareren door een samengevoegde dataset op te bouwen waarin gegevens over vier miljoen productieonderbrekingen gedurende een periode van twee jaar werden gecombineerd met gegevens over verzuim, salariering, personeelsinzet, kennis en kunde van medewerkers, en een berg aan 'harde' and 'zachte' variabelen uit Towers eigen onderneming.

Uit de analyse bleek dat een storing van een lasrobot nauwelijks door de robot zelf werd veroorzaakt. Achtennegentig procent van de variantie in storingstijd bleek terug te voeren op managementvariabelen als salariering, verzuim, overwerkt personeel, en het kennisniveau van de medewerkers aan de assemblageband – het laatste een belangrijke 'zachte' graadmeter die onafhankelijk werd beoordeeld door ter zake deskundige experts met behulp van een standaard dubbelblinde scoringsmethode. De lasrobots van Lamb werden vrijgesproken en de onderneming kon de zaak op een

Met behulp van een samengevoegde dataset werd een conflict opgelost tussen chassisfabrikant Tower en lasrobotfabrikant Lamb

9. J.W. Medcof, 'The Job Characteristics of Computing and Non-Computing Work Activities', *Journal of Occupational and Organizational Psychology*, 69/2, 1996, p. 199-212; V.A. Zeithaml, R.T. Rust en K.N. Lemon, 'The Customer Pyramid: Creating and Serving Profitable Customers', *California Management Review*, 43/4, Zomer 2001, p. 118-146; T.H. Davenport, J.G. Harris, D.W. De Long en A.L. Jacobson, 'Data to Knowledge to Results: Building an Analytic Capability', *California Management Review*, 43/2, Winter 2001, p. 117-138.
10. J.A. Schellenberg, 'The Effect of Pretesting upon the Risky Shift', *Journal of Psychology: Interdisciplinary and Applied*, 88/2, 1974, p. 197-200.
11. W.R. Shadish, T.D. Cook en D.T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin, Boston (Mass.), 2002.

voor haar gunstige manier in der minne schikken. Samengevoegde datasets hebben hun nut ook bewezen bij grote organisaties waar de bedrijfsvoering wordt geconfronteerd met een exponentiële groei in informatie. Walt Disney World gebruikte bijvoorbeeld tot voor kort tientallen losse gegevensbestanden om medewerkers, hotelgasten, restaurantbezoekers en attractiebezoekers te meten. Toen deze bestanden in één alomvattend bestand werden ondergebracht, werd het mogelijk om een rigoureuze analyse uit te voeren op personeelsbehoud, klanttevredenheid en wachttijden bij attracties en restaurants. De analyse liet zien dat een volstrekt onverwachte variabele een voorspeller was voor een gering personeelsverloop: de intrinsieke motivatie (en niet de werknemerstevredenheid, zoals men had verondersteld). Een gering verloop (en niet berichten over 'plezier in het werk', zoals was verondersteld) was vervolgens weer de beste voorspeller voor korte wachttijden bij de restaurants en attracties. Bij Disney loopt nu een voortdurend proces om bestanden samen te voegen, zodat initiatieven op het gebied van personeelsontwikkeling en werving en selectie in één bestand worden gecombineerd waarin informatie is opgeslagen uit diverse voorheen losse aandachtsgebieden.

In deze beide voorbeelden zien we enkele kenmerken (die we in de volgende paragraaf bespreken) die we steeds terug zien bij organisatieonderzoek met behulp van samengevoegde datasets.

Instrument en proces in essentie

In onze benadering van evidence-based management maken we uitgebreid gebruik van standaardmethoden in het organisatieonderzoek. Zoals in al het toegepast onderzoek ligt de nadruk op het gebruik van standaardinstrumenten en herhaalbare processen. Deze gangbare elementen bespreken we hieronder.

Instrument

Het belangrijkste instrument is het centrale gegevensbestand zelf, samengesteld uit diverse zelfstandige rekenbladen die doorgaans in een verschillende frequentie worden geactualiseerd en die worden onderhouden door allerlei verschillende, zelfstandige, onafhankelijke en gedecentraliseerde partijen in de organisatie. Deze decentrale bestanden hebben diverse kenmerken met elkaar gemeen:

- Ze bevatten zowel kwantitatieve als kwalitatieve informatie;
- Ze meten zowel het actuele als het beoogde prestatieniveau voor taken die van doorslagge-

vend belang zijn voor de financiële levensvatbaarheid van de organisatie;

- Ze worden volgens een vast ritme aangevuld (bijvoorbeeld maandelijks), zodat het bestand voortdurend wordt aangevuld met nieuwe gegevens;
- Ze bevatten doorgaans gegevens die verschillen in precisie, objectiviteit en relevantie voor de onderneming;
- Bestanden zijn georganiseerd per bedrijfsonderdeel (bijvoorbeeld divisies) en tijdsperiode;
- Vaak, maar niet altijd, zijn het combinaties van gegevens voor de vier aandachtsgebieden die benoemd zijn in het onderzoek naar de 'balanced scorecard':²⁶ topmanagement (bijvoorbeeld financiële cijfers, productie, kwaliteit); klant (bijvoorbeeld klanttevredenheid, klantbehoud, marktaandeel van de totale bestedingen van de klant [vaak 'share of wallet' genoemd], en klachten); medewerker (bijvoorbeeld personeelsverloop, rechtszaken wegens vermeende discriminatie, klachten, personeelsonderzoek, prestatiebeoordelingen, rapportages over vaardigheden, en gegevens over vrijwillig gevolgde opleidingen); en aandeelhouder (bijvoorbeeld rendement voor de aandeelhouder, deel van de winst beschikbaar voor uitkering aan gewone aandeelhouders (IACS), rendement op het geïnvesteerd vermogen (ROI), en de winst vóór aftrek van rentebetalingen, belasting, afschrijvingen en amortisatie (EBITDA)). Hoewel bij veel ondernemingen een breed scala van gegevens voorhanden is, gebruiken sommige onderzoekers alleen gegevens uit een beperkt aantal aandachtsgebieden van de balanced scorecard, misschien omdat een analyse van de onderlinge relaties tussen graadmeters op een specifiek aandachtsgebied veel inzichten kan opleveren.²⁷

Proces

Organisatieonderzoek met behulp van een samengevoegde dataset is een systematisch, analytisch proces waarbij het prestatievermogen wordt afgemeten aan een objectieve en geschikte norm. In bijna alle gevallen vloeien die normen voort uit een goed gecontroleerde pretest-posttestvergelijking waarbij twee of meer vergelijkbare groepen voor en na een bepaalde interventie met elkaar worden vergeleken. Analyses gebaseerd op pretest-posttestvergelijkingen vormen een cruciaal onderdeel van evidence-based besluitvorming; Pfeffer and Sutton²⁸ zeggen in dit verband ook wel dat de organisatie 'als een prototype' wordt gezien.

12. *Evidence-Based Medicine Working Group, 'Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine', Journal of the American Medical Association, 268/17, 1992, p. 2420-2425; M.L. Green en P.J. Ellis, 'Impact of an Evidence-Based Medicine Curriculum Based on Adult Learning Theory', Journal of General Internal Medicine, 12/12, december 1997, p. 742-750; S.E. Straus en F.A. McAlister, 'Evidence-Based Medicine: A Commentary on Common Criticisms', Canadian Medical Association Journal/Journal De l'Association Médicale Canadienne, 163/7, 2000, p. 837-841.*

13. *Green and Ellis, zie noot 12; R.L. Neame en D.A. Powis, 'Toward Independent Learning: Curricular Design for Assisting Students to Learn How to Learn', Journal of Medical Education, 56/11, november 1981, p. 886-893.*

Pretest-posttestvergelijkingen hebben de volgende kenmerken:

- Ze zijn praktisch, vooral omdat actuele bedrijfsinformatie wordt gemeten zowel voordat een interventie ter beschikking komt van medewerkers of klanten als daarna;
- Ze zijn objectief, ofwel bevoordelen geen groepen in de onderneming;
- Ze zijn niet discreet, zodat het vergelijkingsproces de kerntaak van de onderneming niet verstoort;
- Ze kunnen worden herhaald en vergaand in schaal worden aangepast met het oog op grotere of kleinere delen van de organisatie, omdat standaardisering en consistentie er doorgaans in worden benadrukt;
- Ze zijn methodologisch rigoureuze en verlopen bij voorkeur volgens een gangbaar pretest-posttestontwerp waarbij medewerkers of klanten in een aantal min of meer equivalente groepen worden gecategoriseerd die vervolgens willekeurig worden ingezet voor hetzij het onderzoek (de interventie) dan wel een controlesituatie – waarbij alleen de eerste groep te maken krijgt met de interventie die wordt getest. (Als een pretest geen standaard nulmeting ('baseline') oplevert, wordt in de plaats daarvan soms een benchmark gebruikt.)

Beschrijvende versus diagnostische uitkomsten

Veel ondernemingen beschikken al over een rudimentair analyseprogramma waarbij elementaire 'data mining' wordt gebruikt om een dashboard of balanced scoreboard van belangrijke graadmeters samen te stellen. Met behulp van deze elementaire analyse-instrumenten kunnen veranderingen in de tijd en het actuele prestatieniveau van de onderneming worden gemeten. Ze lenen zich echter slechts voor eenvoudige beschrijvende statistiek (bijvoorbeeld gemiddelden en bandbreedten) en de uitkomsten zijn louter beschrijvend. In de meer geavanceerde toepassingen zijn de data mining en analyse echter veel rigoureuze en de uitkomsten diagnostisch van aard. Deze geavanceerdere onderzoeksprogramma's dienen om de allerdiepste oorzaken te achterhalen, valideringen te maken voor inschattingen naar aanleiding waarvan specifieke initiatieven worden genomen, het effect van interventies te meten, subtiele nieuwe trends in kaart te brengen, prognoses voor het prestatieniveau op te stellen in uiteenlopende scenario's, en herstellende interventies voor te stellen om bepaalde problemen

op te lossen. Om een beeldspraak te gebruiken: het verschil tussen beschrijvende analytische initiatieven en diagnostische analytische initiatieven is als het verschil tussen een patiënt de pols nemen en een compleet diagnostisch onderzoek uitvoeren – echo, röntgen, bloedanalyse, MRI – om een volledig beeld te krijgen van diens actuele medische conditie.

Twee specifieke diagnostische uitkomsten

Het gebruik van een samengevoegde dataset als onderdeel van een doeltreffend evidence-based management levert twee belangrijke diagnostische resultaten op: inzicht in het effect van een bepaalde interventie, en inzicht in het complexe verband tussen twee variabelen (bijvoorbeeld klanttevredenheid en marktaandeel). Deze beide resultaten zijn veel informatiever dan de beschrijvende uitkomsten van een gangbaar dashboard of gangbare scorecard omdat ze de oorzakelijke verbanden tussen variabelen duidelijk maken.

Effectmeting van een bepaalde interventie of een bepaald programma

Hiervoor wordt meestal een pretest-posttestvergelijking uitgevoerd. In de gedragswetenschappen, de volksgezondheid, organisatiepsychologie en dergelijke vakgebieden vormt multivariate inferentiële statistiek (dat wil zeggen, een statistiek met meer dan twee variabelen waarbij *p*-waarden worden gegenereerd, zoals meervoudige regressies, MANOVAs, enz.) een essentieel onderdeel van dergelijke beoordelingen. In het bedrijfsleven, echter, worden soms pretest-posttestvergelijkingen uitgevoerd zonder de statistische tests waarmee onderscheid kan worden gemaakt tussen echte verschillen en verschillen veroorzaakt door een louter toevallige variatie. Deze pretest-posttestvergelijkingen vergen minimaal eenvoudige statistiek zoals correlaties en/of *t*-tests, die beide in gangbare programma's als Microsoft Excel zitten.

Complexe oorzakelijke factor benoemen of meten

Hierbij gaat het erom, oorzakelijke verbanden op te sporen die lastig te meten zijn omdat ze diep verweven zijn met complexe sociale systemen, en ook omdat de analyse ervan veelal in belangrijke mate stoelt op enquêtegegevens en statistische testen. (Het is wel belangrijk om erop te wijzen dat

14. R.B. Briner, 'Improving Stress Assessment: Toward an Evidence-Based Approach to Organizational Stress Interventions', *Journal of Psychosomatic Research*, 43/1, 1997, p. 61-71;
- J. Pfeffer en R.I. Sutton, 'Suppose We Took Evidence-Based Management Seriously: Implications for Reading and Writing Management', *Academy of Management Learning & Education*, 6/1, maart 2007, p. 153-155; J. Pfeffer en R.I. Sutton, *Hard Facts, Dangerous Half-Truths and Total Nonsense: Profiting from Evidence-Based Management*, Harvard Business School, Cambridge (Mass.), 2006.
15. M.D. Flood, 'Embracing Change: Financial Informatics and Risk Analytics', *Quantitative Finance*, 9/3, 2009, p. 243-256; T.H. Davenport, 'Competing on Analytics', *Harvard Business Review*, 84/1, januari 2006, p. 99-107; Davenport c.s., zie noot 9.

Een samengevoegde dataset werpt licht op oorzakelijke verbanden tussen variabelen – dat is veel informatiever dan een gebruikelijke scorecard

16. W. Yeoh, A. Koronios en J. Gao, 'Managing the Implementation of Business Intelligence Systems: A Critical Success Factors Framework', *International Journal of Enterprise Information Systems*, 4/3, 2008, p. 79-95.

17. E.M. Kyrrouz en K. Humphreys, 'Do Health Care Workplaces Affect Treatment Environments?', *Journal of Community & Applied Social Psychology*, 7/2, april 1997, p. 105-118.

18. D.E. Bowen en C. Ostroff, 'Understanding HRM-Firm Performance Linkages: The Role of the "Strength" of the HRM System', *Academy of Management Review*, 29/2, april 2004, p. 203-221.

19. R. Waldersee en F. Luthans, 'The Impact of Positive and Corrective Feedback on Customer Service Performance', *Journal of Organizational Behavior*, 15/1, januari 1994, p. 83-95.

20. R.P. Vecchio, 'The Impact of Referral Sources on Employee Attitudes: Evidence from a National Sample', *Journal of Management Studies*, 21/5, 1995, p. 953-965.

veel van de geavanceerde statistische instrumenten die nodig zijn voor het analyseren van complexe verbanden, zoals meervoudige regressie, nu voor het grijpen zijn in Microsoft Excel.) Complexe oorzakelijke factoren ('drivers') zijn gebruikelijk in organisatieonderzoek: vermengende variabelen ('confounding variables'; dat wil zeggen, ongemeten factoren die van invloed zijn op één of meer kritieke variabelen in het model) die de interpretatie kunnen verstoren; bemiddelende variabelen ('mediating variables'; dat wil zeggen, factoren tussen twee belangrijke variabelen in, zoals in een inkomensmodel kunde tussen jarenlange opleiding en salaris kan liggen); modererende variabelen ('moderating variables'; dat wil zeggen, factoren die de invloed van de ene op een andere variabele drastisch veranderen, zoals het geslacht van invloed kan zijn op het effect van lichaamsbeweging op de gezondheid); en breed gespreide effecten die een diffuus kenmerk kunnen vormen (zoals een nadruk op persoonlijke verantwoordelijkheid), waarvan de bedrijfscultuur doordrenkt lijkt te zijn. In toegepast onderzoek waarbij samengevoegde datasets worden gebruikt om complexe oorzakelijke factoren te analyseren, worden meestal aspecten als de volgende gemeten: de grootte van een bekend sterk verband (bijvoorbeeld tussen de betrokkenheid van medewerkers en het personeelsverloop); het effect van een zwak en slecht begrepen verband (bijvoorbeeld tussen teamwerk en productiefouten); een zwak verband dat geworteld is in een complex systeem (bijvoorbeeld ROI of bonussen om medewerkers vast te houden); een verband dat recentelijk sterker is geworden (bijvoorbeeld van een onlangs uitgebreid programma voor de klantenservice); of een zwak verband waarbij een uitkomst juridische consequenties kan hebben (bijvoorbeeld omdat er sprake zou kunnen blijken te zijn van discriminatie). Voor bepaalde statistische analyses zijn specialisten nodig, maar in het algemeen zijn door de toegenomen beschikbaarheid van statistische instrumenten op desktopcomputers de mogelijkheden voor organisatieonderzoek van een behoorlijke omvang en waarde in het afgelopen decennium ongetwijfeld sterk toegenomen. Beide diagnostische uitkomsten stellen de onderzoeker voor een netelig probleem: het is vaak buitengewoon moeilijk om mensen in het bedrijfsleven die niet thuis zijn in statistiek te overtuigen van de omvang en het gewicht van een oorzakelijk verband. Een deel van de doelgroep lijkt al gauw ten onrechte te denken dat elk plausibel effect groot en universeel is, en een ander deel lijkt zijn

scepsis over het nut van ieder kwantitatief onderzoek maar niet te kunnen overwinnen. Wij vermoeden dat dit komt doordat onderzoekers te veel geneigd zijn om snel bewijzen aan te dragen en zich onvoldoende afvragen wat er nu precies voor nodig is om een oorzakelijk verband te kunnen aantonen. Zeker in het bedrijfsleven worden hun ogenschijnlijk oprechte argumenten in de ogen van de doelgroep vaak ondermijnd doordat er te stellige beweringen worden gedaan of doordat er te simplistische, misleidende of vertekende grafieken worden getoond.

Met een doordachte voorbereiding en aandacht voor detail kunnen evidence-based programma's gestoeld op samengevoegde datasets (en evenzeer overtuigende grafieken) goed gedocumenteerde en gepast voorzichtige argumenten opleveren; daarmee kan ook een doelgroep die geen bijzondere affiniteit heeft met statistiek of kwantitatieve analyse worden overtuigd van de geconstateerde oorzakelijke verbanden.

Enkele toonaangevende onderzoeken

Bij de onderzoeken die we hier bij wijze van voorbeeld willen bespreken, draait het om drie cruciale vragen: wat wordt er gemeten, hoe kunnen we de uitkomsten controleren om niet ter zake doende factoren uit te sluiten, en wanneer dient de pretest-posttestvergelijking te worden uitgevoerd? Aangezien alle mogelijke antwoorden op deze vragen in ons geval tweeledig zijn, hebben we acht mogelijke onderzoekscategorieën. We meten hetzij een programma dan wel een 'zachte' oorzakelijke factor die een belangrijk aspect van de bedrijfscultuur aangeeft, zoals leiderschap, normen en waarden of communicatie; we zorgen voor de benodigde controle hetzij door een willekeurige controlegroep aan te wijzen die niet aan de te meten interventie wordt onderworpen, dan wel door middel van een statistische controlevariabele (in de vorm van een co-variabele) op mogelijk vermengende factoren; en we organiseren pretest-posttestvergelijkingen, hetzij volgens een waternivauschema (waarbij interventies en metingen voortdurend op verschillende momenten plaatsvinden), dan wel volgens een klassiek pretest-posttestontwerp met een willekeurig aangewezen onderzoeksgroep waarbij alle pretests tegelijkertijd worden uitgevoerd, alle interventies tegelijkertijd plaatsvinden, en alle posttests tegelijkertijd worden gedaan. Omwille van de beknoptheid zullen in de nu volgende voorbeelden alleen de vier meest voorkomende van deze acht combinaties worden beschreven.

Voorbeeld 1: effectmeting van een verbeteringsprogramma met behulp van een willekeurig gecontroleerde proef

Bij de Panda Restaurant Group hebben we het effect van de elektronische studiemodules van Root Learning getest. Dit zijn lesprogramma's op de computer over winstgevendheid, productiviteit en klanttevredenheid die de cursist in zijn/haar eigen tempo kan doorlopen. We kozen voor een klassiek pretest-posttestontwerp met een onderzoeksgroep en een groep die geen interventie onderging. De groepen werden willekeurig aangewezen. Om precies te zijn: restaurants van de keten werden willekeurig ingedeeld in de onderzoeks- dan wel de controlegroep. We zorgden ervoor dat elk van de beide groepen een vergelijkbaar deel had van de verschillende locaties (bijvoorbeeld in een winkelcentrum versus vrijstaand) en sub-locaties (bijvoorbeeld direct aan de straat of binnenin). De onderzoeksgroep en de controlegroep bestonden elk uit zestien restaurants. De restaurants in de onderzoeksgroep kregen de beschikking over de elektronische studiemodules ('e-learning modules') ontwikkeld door Root, de restaurants in de controlegroep niet. De analyse werd uitgevoerd volgens een klassiek pretest-posttestontwerp met willekeurige toewijzing. Dat wil zeggen, de onderzoeksgroep en de controlegroep werden willekeurig aangewezen en beide werden twee keer, en tegelijkertijd, gemeten – een keer voordat de interventie plaatsvond, en een keer nadat de interventie was afgerond.

Het onderzoek werd onverwacht verstoord door de extra aandacht die sommige restaurants in zowel de onderzoeks- als de controlegroep kregen toen een aantal hooggeplaatste managers en directieleden tijdens het onderzoek onverwacht enkele vestigingen bezochten. Toch lijkt uit de einduitkomsten te kunnen worden afgeleid dat deze bezoeken slechts een beperkte en tijdelijke prestatieverbetering te weeg brachten in de restaurants, een bevinding die aansluit met behoorlijk wat gepubliceerd onderzoek.

Het pretest-posttestonderdeel van de analyse was cruciaal. Dit hield in dat het prestatieniveau van de beide groepen restaurants zowel voor als na de interventieperiode op één en hetzelfde moment werd beoordeeld aan de hand van identieke graadmeters. In dit geval vond de meting plaats aan de hand van een scala van graadmeters voor respectievelijk de winst, het aantal verkooptransacties, klanttevredenheid en tal van andere belangrijke variabelen. Een voorafgaand onderdeel van de analyse – en in wezen het onderdeel waardoor een dergelijk on-

derzoeksontwerp nuttige informatie oplevert – was een evaluatie van de seizoenstrends. In de maand waarin de modules voor e-learning ter beschikking werden gesteld van de onderzoeksgroep, gaven de duizend restaurants van Panda-keten (gemiddeld) een daling te zien in de bruto omzet en alle belangrijke kengetallen voor de productiviteit terwijl de klanttevredenheid steeg. De 32 restaurants in de interventie- en de controlegroep hadden ongetwijfeld te maken met precies dezelfde seizoensfactoren, maar in de deze beide groepen steeg het prestatieniveau enigszins gedurende de proefperiode. Wellicht werd deze lichte prestatieverbetering ten dele veroorzaakt door het Hawthorne-effect, dat wil zeggen dat het prestatievermogen gestegen zou zijn door de extra aandacht die medewerkers kregen doordat het management langs was gekomen. Het bescheiden effect van de seizoenstrends en de managementbezoeken aan de vestigingen daargelaten, was de hoofdbevinding uit het onderzoek zonneklaar: het prestatieniveau steeg meer bij de restaurants in de onderzoeksgroep dan bij de restaurants in de controlegroep, en het ging om een zowel economisch als statistisch significant verschil. Enkele bevindingen in detail: het aantal verkooptransacties steeg in beide groepen, maar het aantal transacties steeg significant meer in de restaurants waar medewerkers modules voor e-learning gebruikten. Ook de bruto omzet steeg in beide groepen, maar significant meer in de restaurants waar de medewerkers met de modules voor e-learning werkten. Idem voor de productiviteitscijfers: ook die stegen in beide groepen, maar de totale productiviteitsstijging was significant groter in de restaurants waar de modules voor e-learning waren ingezet. Opmerkelijk was dat de restaurants van de onderzoeksgroep niet alleen aanzienlijk meer omzet maakten en hun winstgevendheid en productiviteit verbeterden, maar ook de klanttevredenheid zagen stijgen, en bovendien naar een significant hoger niveau dan bij de restaurants in de controlegroep.

Uit de uitkomsten bleek duidelijk, consistent en overtuigend dat de restaurants in de onderzoeksgroep beter hadden gepresteerd dan in de controlegroep. Omdat de beide groepen louter willekeurig waren samengesteld, kunnen we er vrij zeker van zijn dat het gebruik van de modules voor e-learning van Root de beste verklaring biedt voor de aanhoudende en algemene prestatieverbetering onder de restaurants in de onderzoeksgroep.

21. J. Barling, T. Weber en E.K. Kelloway, 'Effects of Transformational Leadership Training on Attitudinal and Financial Outcomes: A Field Experiment', *Journal of Applied Psychology*, 81/6, 1996, p. 827-32.
22. B.S. Maini en P. Morrel-Samuels, 'Cascading Improvements in Communication: Adopting a New Approach to Organizational Communication', *Physician Executive*, 32/5, september/oktober 2006, p. 38-43.
23. R.W. Eichinger en M.M. Lombardo, 'The ROI on People – The 7 Vectors of Research', niet uitgegeven manuscript, verspreid door Lominger Limited LTD, Minneapolis, 2004.
24. SAS Institute, *JMP Statistics and Graphics Guide*, SAS Institute, Inc., Cary (Noord-Carolina), 2007.
25. J. Pfeffer en A. Davis-Blake, 'Unions and Job Satisfaction: An Alternative View', *Work and Occupations*, 17/3, 1990, p. 259-283.
26. R.S. Kaplan en D.P. Norton, 'The Balanced Scorecard; Measures that Drive Performance', *Harvard Business Review*, 70/1, januari/februari 1992, p. 71-80.

Voorbeeld 2: effectmeting van een verbeteringsprogramma met behulp van een niet willekeurig gekozen vergelijkingsgroep

Een willekeurig aangewezen controlegroep is de norm in quasi-experimenteel medisch onderzoek. Bij onderzoek in een organisatorische context worden veelal statistische controlevariabelen gebruikt (in plaats van een willekeurig toegepaste experimentele ingreep), een aanpak die in onze waarneming goed heeft uitgedaan bij honderden ondernemingen. Enkele korte voorbeelden volstaan.

In een typerend voorbeeld vroeg EDS aan Root om een reeks *train the trainer*-discussiegroepen te verzorgen waarbij gebruik werd gemaakt van een aantal Learning Map®-modules. Het doel was om de zestigduizend medewerkers van EDS-GM te helpen, de klanttevredenheid te verbeteren. In het kader van een ander contract, dat gelijktijdig werd gesloten maar daar los van stond, vroeg EDS aan EMPA (het bedrijf van de eerste auteur) om een objectieve programma-evaluatie te verzorgen waaruit zou blijken of de interventie van Root het gewenste effect sorteerde. EMPA gebruikte een standaardontwerp voor een pretest-posttestvergelijking met statistische controlevariabelen (bijvoorbeeld op het effect van een wisselvallige deelname aan het programma), zodat enquêtegegevens en objectieve prestatiemetingen op twee momenten met elkaar konden worden vergeleken, namelijk de periode waarin het programma werd uitgevoerd (de interventie) en zes maanden later. In de posttest-enquête werd iedere deelnemer expliciet gevraagd of zij de Learning Map®-discussies hadden bijgewoond en in de zes maanden daarna de methoden, die in die bijeenkomsten waren aanbevolen, hadden toegepast.

Met behulp van deze statistische controlevariabelen konden we de medewerkers die hadden deelgenomen aan de discussiebijeenkomsten vergelijken met de medewerkers die er niet aan hadden deelgenomen. Nog belangrijker, echter, was dat we onderscheid konden maken tussen degenen die de inzichten, die in de discussiegroepen waren aangedragen, wel respectievelijk niet waren gaan toepassen in hun werk.

Deze statistische controlevariabele fungeerde als een cruciale co-variabele die de deelnemers aan de trainingen onderscheidde van de medewerkers die een manier hadden gevonden om zich eraan te onttrekken (hoewel de deelname verplicht was). Een probleem is natuurlijk dat er allerlei redenen kunnen zijn waarom iemand niet aan zo'n training meedoet. Wellicht vonden sommigen dat zij al een

uitstekende klantenservice leverden. Of misschien waren sommigen per definitie sceptisch over nieuwe initiatieven die de onderneming lanceerde. Of wellicht schatte men de klanttevredenheid bij het eigen bedrijfsonderdeel dermate laag in, dat er met de gebruikelijke verbeteringsmethoden niets meer aan te doen viel zolang er geen vergaande veranderingen werden doorgevoerd in de producten of de servicegaranties.

De cruciale vergelijking betrof in dit geval niet degenen die wel respectievelijk niet hadden deelgenomen, maar de deelnemers die de aangedragen inzichten wel respectievelijk niet waren gaan toepassen in hun werk. Een statistische analyse toonde aan dat het zin had om de aangedragen methode toe te passen:

- Bij medewerkers die niet hadden deelgenomen aan de discussies werd naderhand ook geen verbetering in de klanttevredenheid gemeten;
- Ook bij medewerkers die wel aan de discussies hadden deelgenomen, maar die de daarin aangedragen aanpak niet hadden toegepast, steeg de scores voor klanttevredenheid vervolgens niet;
- Bij degenen die wel hadden meegedaan aan de Learning Map®-discussiegroepen en die vervolgens ook aan de slag waren gegaan met de voorgestelde aanpak, steeg de klanttevredenheid gedurende de zes maanden tussen de pretest- en de posttest-enquête met tien procent.

Bovendien bleek dat, naarmate de medewerkers de aangereikte informatie meer gingen gebruiken, ook de klanttevredenheid verder steeg. Een dergelijke uitkomst wordt zelden geconstateerd wanneer een oorzakelijk verband ontbreekt. De uitkomsten illustreren al met al vrij goed hoe de quasi-experimentele benadering kan worden gebruikt om de uitkomsten te controleren op vermengende variabelen wanneer het niet mogelijk is om een willekeurige controlegroep aan te wijzen (zie ook kader 'Inleiding tot quasi-experimentele analyse').

Voorbeeld 3: effectmeting van een oorzakelijke factor ('driver') met behulp van een statistische controle

Dezelfde aanpak als hierboven beschreven kan worden gebruikt om het effect te meten van complexe oorzakelijke factoren ('drivers') in de bedrijfscultuur, zoals teamwerk, klantgerichtheid of intrinsieke motivatie. Zo heeft Fallon Clinic – de grootste groepspraktijk in centraal Massachusetts – vier jaar achter elkaar een enquête gehouden onder het voltallige personeel in de 34 klinieken van de

27. Dank aan een van onze reviewers, die dit punt noemde; het bewijst maar weer eens dat er veel over een organisatie te leren valt, zelfs wanneer er niet volgens het boekje een uitgebreid theoretisch kader wordt aangehouden.

28. J. Pfeffer en R.I. Sutton, 'Treat Your Organization as a Prototype: The Essence of Evidence-Based Management', *Design Management Review*, 17/3, zomer 2006, p. 10-14.

Inleiding tot quasi-experimentele analyse

Quasi-experimentele analyse is in de vroege jaren zestig ontwikkeld door onderzoekerspsychologen die enkele methoden wilden formuleren voor onderzoek buiten het laboratorium, waarbij geen strikte experimentele controle vereist was.* Quasi-experimentele analyse is tegenwoordig één van de belangrijkste analytische methoden voor klinisch onderzoekers in de geneeskunde, kwantitatief georiënteerde organisatiepsychologen, onderzoekers die op grond van evidence-based methoden tot beslissingen komen, en getuigendeskundigen die statistisch bewijs van discriminatie op de werkplek onderzoeken; dit zijn slechts enkele voorbeelden.

Quasi-experimentele analyse biedt twee belangrijke alternatieven voor het vormen van controlegroepen: een willekeurig aangewezen controlegroep die niet aan de behandeling wordt onderworpen – doorgaans de experimentele controle genoemd; of een aantal statistische controlevariabelen op mogelijk vermengende factoren die zich niet gemakkelijk lenen voor experimentele controle (zoals lichaamsgewicht of bloeddruk). Er is een cruciaal verschil tussen een (niet aan de interventie onderworpen) controlegroep bij een experimentele ingreep en statistische controle. In het eerste geval gaat het om dubbelblind tests waarbij placebo's worden gebruikt in plaats van de medicatie waar het allemaal om begonnen is. Het tweede is gebruikelijk in klinisch en organisatorisch onderzoek waarbij bepaalde variabelen, zoals het aantal jaren genoten onderwijs of het aantal personeelsleden, niet worden veranderd maar slechts worden gemeten als een co-variable die de statisticus er bij de analyse uit haalt.

De onderzoeker die een quasi-experimentele analyse uitvoert, meet doorgaans effecten en traceert de diepste oorzaken door een pretest-posttestvergelijking uit te voeren op grond van één van twee mogelijke schema's: een klassiek pretest-posttestschema met willekeurige aanwijzing – waarbij de onderzoeksgroep en de controlegroep tegelijkertijd worden getest, zowel voor als na de interventie – of een cyclisch schema, waarbij een groep op gezette tijden (bijvoorbeeld maandelijks of jaarlijks) met zichzelf en met vergelijkbare groepen ('peer groups') wordt vergeleken.

Een eenvoudige vergelijking kan in dezen misschien verhelderend werken: een quasi-experimentele analyse is voor gegevens wat tijdregistratie is voor een polshorloge. Een polshorloge is handig omdat mensen er dezelfde methode op na houden om de tijd af te lezen in minuten, seconden en uren. Door deze methode toe te passen, kunnen we een aantal gewenste uitkomsten verkrijgen. Het elementaire instrument, de methode, en de uitkomsten zijn alle gerelateerde maar aparte eenheden. Om de vergelijking door te trekken: de meeste ondernemingen gebruiken analyses vooralsnog met zo weinig inzicht dat je het zou kunnen vergelijken met een klok die uitsluitend wordt gebruikt om de huidige tijd af te lezen – zonder enig besef dat met die zelfde klok ook plannen op elkaar kunnen worden afgestemd, de tijdsduur kan worden gemeten, en verandering kan worden geregistreerd.

De lezer die meer over dit onderwerp zou willen weten, wordt verwezen naar de klassieke teksten erover van T.D. Cook en D.T. Campbell* of hun collega W.R. Shadish.**

* T.D. Cook en D.T. Campbell, *Quasi Experimentation: Design and Analysis Issues for Field Settings*, Rand McNally, Chicago (Ill.), 1979.

** W.R. Shadish, T.D. Cook en D.T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin, Boston (Mass.), 2002.

groep. Elk jaar opnieuw bleek uit een statistische analyse van de samengevoegde dataset dat communicatie tussen de medewerkers van grote invloed was op de patiënttevredenheid en de winstgevendheid. Zoals geschetst in een recent artikel in *Physician Executive*,²⁹ bleek uit de uitkomsten van de enquêtes dat de initiatieven van de ceo om de communicatie in en door de gehele organisatie te veranderen en te verbeteren, inderdaad het gewenste effect hadden: de communicatie verbeterde elk jaar van het onderzoek.

Belangrijker, en waardevoller vanuit financieel perspectief, was dat deze verbeteringen in de commu-

nicatie aanvullende voordelen leken op te leveren, zoals een grotere patiënttevredenheid en hogere winstgevendheid. (De uitkomsten voor de winstgevendheid en de patiënttevredenheid waren vergelijkbaar; in deze korte beschrijving beperken wij ons tot de financiële prestatie.) In onze analyse van de samengevoegde dataset voor alle jaren van 2002 tot en met 2005, constateerden wij alle vier de kenmerken van een oorzakelijk verband tussen communicatie en winstgevendheid: communicatie heeft een gelijktijdige associatie met winstgevendheid; communicatie in het lopende jaar heeft voorspellende waarde voor de winstgevendheid van

29. Maini en Morrel-Samuels, zie noot 22.

het volgende jaar; geen enkele andere beschikbare 'zachte' variabele die wij maten – zoals waargenomen teamwerk, leiderschap, bedrijfsmiddelen, salariëring en secundaire arbeidsvoorwaarden, of eerlijkheid – vertoonde een statistisch significant verband met de winstgevendheid; hoe meer de communicatie van het ene op het andere jaar verbeterde, des te meer steeg ook de winst van het ene op het volgende jaar. Uit archiefonderzoek blijkt dat wanneer deze vier kenmerken – associatie, voorspellende waarde, uitsluiting van alternatieven en dosisafhankelijkheid – gezamenlijk worden gepresenteerd, dit kennelijk volstaat om mensen zonder een achtergrond in statistiek die in een rechts-

Vaak zal de doelgroep niet onderlegd zijn in statistiek – een doordachte presentatie is daarom cruciaal

zaak een oordeel moeten vellen (lees: rechters en juryleden) ervan te overtuigen dat er inderdaad sprake is van een causaal verband.³⁰ Aangezien juridische normen vermoedelijk stringenter zijn dan de meeste informele criteria, mogen we er gevoeglijk op vertrouwen dat een dergelijk op causaliteit duidend bewijs tamelijk overtuigend zal zijn in een onderneming, zelfs al is er geen absolute zekerheid te geven voor een oorzakelijk verband – zoals dat nu eenmaal in nagenoeg ieder onderzoek geldt voor de interpretatie van de uitkomsten.

Het is interessant om erop te wijzen dat communicatie ook in veel van onze andere analyses een onverwacht belangrijke oorzakelijke factor bleek te zijn. Zo hebben bijvoorbeeld de Arthur W. Page Society, een organisatie van Amerikaanse topmanagers voor communicatie, en Gagen MacDonald, een communicatieadviesbureau in Chicago, een onderzoek laten uitvoeren naar het voorspellende effect van communicatie op de aandelenkoers. Uit een klassieke pretest-posttestvergelijking onder de twintig deelnemende ondernemingen bleek dat, naarmate de communicatie in de onderneming beter was, het financiële rendement voor de aandeelhouder hoger uitviel. In deze bevindingen klinken de resultaten door van de National Benchmark Study,³¹ waarin de relatie tussen de bedrijfscultuur en de daaropvolgende financiële resultaten landelijk werden gemeten.

Voorbeeld 4: Effectmeting van een programma waarbij vroege en late gebruikers met elkaar worden vergeleken
Recentelijk hebben wij een grote effectanalyse ('impact analysis') afgerond bij Holiday Inn Express (HiEx), waar we het effect van de e-learningmodules van Root Learning hebben gemeten

aan de hand van twee cruciale maatstaven in de hotelaanpak: omzet per beschikbare kamer ('Revenue Per Available Room', afgekort RevPAR) en het aantal klachten van gasten. De analyse moest onder bijzonder lastige omstandigheden worden uitgevoerd. Om te beginnen werkt de moedermaatschappij Inter-Continental Hotels Group, met franchisecontracten die de directeur-eigenaar van een HiEx-vestiging veel zelfstandigheid en privacy bieden. Dat betekent dat het hoofdkantoor niet kan beschikken over veel van de dagelijkse graadmeters (zoals het aantal personeelsleden) die doorgaans belangrijk zijn bij quasi-experimentele analyse in een onderneming.

Nog problematischer was dat de e-learning modules volgens een geleidelijk introductieschema werden ingevoerd. De afzonderlijke hotels maakten er in verschillende mate en op verschillende tijden gebruik van; de medewerkers van de afzonderlijke hotels konden net zoveel of net zo weinig van de e-learning modules gebruiken als zij zelf (en vermoedelijk hun directeuren) wilden. Daar kwam nog bij dat HiEx iedere dag ongeveer drie kenmerken toevoegde aan zijn groep van 1.600 hotels. Al met al leek de organisatie veel te fluïde voor welke gedetailleerde analyse dan ook. Het meetsysteem ('data tracking system') bij HiEx is echter een toonbeeld van efficiëntie en precisie, waardoor wij de ontbrekende en veranderende gegevens konden compenseren door belangrijke co-variabelen te gebruiken, zoals het aantal kamers, merk en type locatie. Aan het eind van het onderzoek, dat een half jaar in beslag nam, hadden we voor elk HiEx-hotel over de hele wereld een schone set met graadmeters voor het prestatievermogen.

We voegden de prestatiegegevens voor de hotels samen met de gegevens van Root Learning voor de dertig e-learning modules die zij aan HiEx leverden. De gegevens waren inclusief vijf identieke pretest- en posttestvragen over de inhoud van iedere module. Met een eenvoudige meerkeuzetoets konden we op een rudimentaire maar duidelijke manier de kennis van een medewerker meten vóór en nadat deze iedere e-learning module deed. Een voorafgaande statistische analyse gaf een normale verdeling van deze kennisscores onder de medewerkers te zien, alsmede significante verschillen in de scores tussen de hotels onderling – belangrijke voorwaarden voor een uitgebreide statistische analyse omdat ze suggereren dat de scores voor de medewerkers bijzonder waardevol en nuttig voor de onderneming zijn.

30. P. Morrel-Samuels and P.J. Jacobson, 'Using Statistical Evidence to Prove Causality to Non-Statisticians', voordracht aan de Meeting of the American Psychology-Law Society, Jacksonville (Florida), 2008.

31. P. Morrel-Samuels, 'The National Benchmark Study: Employee Motivation Affects Subsequent Stock Price', voordracht aan de Meeting of the American Psychological Association, Toronto (Canada), augustus 2009.

We gebruikten samengevoegde datasets om te meten welk effect de e-learning modules hadden naarmate ze in het hele systeem werden geïntroduceerd. Dat leverde overtuigend bewijs op dat het gebruik van de e-learning modules de omzet bevorderde en klachten van gasten deed vermindere. Om precies te zijn, de kennisscores waren significant gecorreleerd met de gelijktijdige omzet per beschikbare kamer, de kennisscores hadden voorspellende waarde voor de omzet per beschikbare kamer in de volgende maand, en hoe groter de stijging in de kennisscore in de huidige maand, des te groter de stijging in de omzet per beschikbare kamer in de volgende maand.

We kregen bovendien nagenoeg identieke uitkomsten toen we de relatie analyseerde tussen de kennisscores en klachten van gasten. Kennisscores hielden verband met een afname van de klachten van gasten in dezelfde maand, waren een significante voorspeller van minder klachten in de volgende maand, en boden ook duidelijk bewijs van voorspellende dosisafhankelijkheid: hoe meer de kennisscore in de lopende maand steeg, des te sterker daalde in de volgende maand het aantal klachten.

Een cruciaal aspect van het evidence-based managementprogramma bij HiEx was een sequentieel introductieschema in combinatie met statistische controle. Dat wil zeggen, het gebruik van nieuwe e-learning modules werd gedurende de gehele onderzoeksperiode in elk hotel bepaald door de dagelijkse vrije keuze van de medewerkers. Dit betekende dat ons onderzoeksontwerp voorzag in een statistische controlevariabele voor elk hotel in de onderneming. Hierdoor kon iedere maand precies worden gemeten hoe de kennisscores per hotel samenhangen met het objectief gemeten prestatieniveau van dat zelfde hotel. Omdat de gegevens bovendien zowel terug als vooruit in de tijd werden aangevuld konden er maandelijkse veranderingsscores worden gegenereerd, en kon het maandelijkse prestatieniveau worden geanalyseerd onder statistische controle op het prestatieniveau van de voorafgaande maand. Door de samengevoegde dataset en het quasi-experimentele ontwerp konden duizenden rigoureuze pretest-posttestvergelijkingen worden uitgevoerd met een eenvoudige multivariate statistische test waarbij de veranderingsscores voor de omzet werden onderzocht als een functie van veranderingsscores voor kennis (met een controle op mogelijke vermengende effecten die verband hielden met de omvang van het hotel, het seizoen, de locatie en tal van andere factoren).

Wel is het zo dat de uitkomsten van het onderzoek werden afgezwakt door het gegeven dat wij – zoals veel onderzoekers, om redenen die Florida and Davison hebben samengevat³² – niet het gezag hadden om willekeurig geselecteerde hoteleigenaren de e-learning modules al dan niet te laten gebruiken. Een dergelijke zeggenschap had statistisch gezien allerlei voordelen opgeleverd³³ en een controle geboden op nog meer vormen van vermenging, zoals leergerigheid of het aantal afleidingen op de werkplek. Dat neemt echter niet weg dat hoe meer kennis de medewerkers opdeden, des te meer de omzet per beschikbare kamer vervolgens steeg – overtuigend bewijs van een zeker oorzakelijk verband, al is het dan met een gepaste waarschuwing. Al met al laat dit voorbeeld heel redelijk zien dat er ondanks een vrij geringe zeggenschap over de omstandigheden van een experiment toch heel wat inzichten kunnen worden verkregen.

Waarom evidence-based management soms mislukt

Ondernemingen lanceren een evidence-based programma soms inderhaast en zonder de benodigde weloverwogen voorbereiding. Daarom is het zinloos om bij wijze van waarschuwing enkele voorbeelden te geven van initiatieven die mislukt zijn, en om de zes belangrijkste obstakels voor een succesvol evidence-based managementprogramma te noemen.

Soms klopt alles aan een evidence-based initiatief behalve de uitkomsten. Zo liet bijvoorbeeld één van de grootste Amerikaanse autofabrikanten (een onderneming die nu overeind probeert te krabbelen na een faillissement) enkele jaren geleden een rigoureuze onderzoek uitvoeren naar de vaardigheden van zijn 35.000 technici in de VS. Men schrok van de uitkomsten. Het onderzoek naar vaardigheden liet duidelijk zien dat bepaalde groepen technici hun eigen vaardigheidsniveau consequent te hoog inschatten, te weinig gebruikmaakten van de vrijwillige trainingsprogramma's, en productielijnen ontwierpen die leidden tot veel gebreken aan de producten, slechte klanttevredenheidsscores en matige verkoopcijfers. Toen de volledige resultaten van de statistische analyse beschikbaar waren, vertelde de projectmanager zijn team niet zonder schaamte dat het, gezien de aversie tegen verandering in de onderneming, gemakkelijker was om maar wat oppervlakkige veranderingen door

Hoe beter de communicatie in de onderneming, des te hoger het financiële rendement voor de aandeelhouder

32. R. Florida and D. Davison, 'Gaining from Green Management: Environmental Management Systems Inside and Outside the Factory', *California Management Review*, 43/3, Lente 2001, p. 64-86.

33. S.B. Morris, 'Estimating Effect Sizes from Pretest-Posttest-Control Group Designs', *Organizational Research Methods*, 11/2, 2007, p. 364-386.

te voeren in het trainingsaanbod voor de technici en het project voor geslaagd te verklaren dan om het onderliggende probleem aan te pakken, met aanvullende training die wel aankwam en door medewerkers over te plaatsen of taakverantwoordelijkheden te veranderen.

Een ander voorbeeld was de ambitieuze maar niet goed van tevoren uitgewerkte implementatie van een analyseprogramma bij één van Amerika's grootste en meest winstgevendende banken. Het programma moest degelijk bewijs vergaren ter ondersteuning van alle stadia van talentmanagement voor de kwart miljoen medewerkers. Het aantal promoties en overplaatsingen en het volume aan gegevens was echter zo groot dat de analisten belangrijke onderscheiden uit het oog verloren. Bijvoorbeeld: een gewone managementbenoeming versus een benoeming van een ervaren manager bij een kwakkelend bedrijfsonderdeel dat uit het slop getrokken moest worden. Doordat dergelijke situaties over één kam werden geschoren, werden de saneringsmanagers onterecht gestraft voor de slechte financiële resultaten en uitkomsten uit de personeelsenquête van hun bedrijfsonderdeel. Uiteindelijk verloor dit analyse-initiatief ieder nut voor de onderneming

en werd het gestaakt omdat de datapunten niet in hun juiste en volledige context werden gepresenteerd.

Evidence-based management ontspoorde in deze twee voorbeelden doordat men niet goed begreep hoe een aangetoond oorzakelijk verband in organisatieonderzoek moet worden geïnterpreteerd. Vermoedelijk was er overigens ook sprake van een gebrek aan moed in het projectteam. Deze problemen werden verergerd door de sterke verkokering in de organisatie, een gebrek aan steun van het topmanagement, verzet tegen verandering, conflicten tussen verantwoordelijkheidsgebieden, en, bovenal, een wijd verspreid misverstand omtrent de methoden en doelen van evidence-based management.

Zeven mogelijke obstakels voor succes

Er is geen bijzonder statistisch instrument voor nodig om een samengevoegde dataset te maken of te analyseren. Initiatieven voor evidence-based management lopen dan ook niet stuk op een gebrek aan geavanceerde statistische analyse. Volgens ons zijn verzet tegen verandering en een gebrekkig inzicht de grootste obstakels voor een succesvolle implementatie. Onze ervaring leert echter dat een

degelijk programma voor evidence-based management zelf al – tot op zekere hoogte – kan helpen om het verzet tegen verandering in de organisatie te slechten en het gebrek aan inzicht in en buiten het projectteam te verhelpen. Een werkelijk succesvolle implementatie vereist echter dat ook de volgende obstakels worden overwonnen:

Meting zonder validiteit of nut

Als een belangrijke meting (bijvoorbeeld een enquête voor klantonderzoek) niet valide is, dan kunnen de houding, perceptie, problemen en wensen van de respondenten niet nauwkeurig worden gemeten. De daaropvolgende interventies om de respondenten te helpen, zijn dan gebaseerd op misleidende gegevens. (Zie 'Getting the Truth into Workplace Surveys'³⁴ voor nadere informatie over het belang van validiteit en zakelijk nut voor onderzoeksgegevens.)

Gebrek aan aandacht voor de motivatie van de medewerkers

Als de motivatie van de medewerkers (en vergelijkbare begrippen als betrokkenheid, commitment of loyaliteit) geen belangrijk aandachtspunt is in een onderneming, dan gaan de medewerkers op hun werkplek 'met pensioen', zoals een van onze cliënten het recentelijk formuleerde. Geen enkel initiatief en geen enkele verbetering op personeelsgebied – ook al zijn die nog zozeer gebaseerd op overtuigend bewijs – zal dan merkbare invloed kunnen uitoefenen op het prestatievermogen van de onderneming. Dit is een belangrijk aandachtspunt omdat buitensporig cynisme onder het personeel funest is voor evidence-based management: medewerkers moeten zich willen verbeteren; zonder die essentiële motivatie loopt ieder initiatief stuk op de rotsen van cynisme en onverschilligheid, ongeacht hoe goed het betreffende evidence-based programma ook moge zijn.

Slechte samenwerking en communicatie

Als de communicatie in de organisatie mank gaat aan gebreken en obstakels die men niet wil erkennen, en aan chronische onderlinge conflicten, dan ondermijnt dat de coördinatie tussen afdelingen en mensen. Dat gaat bijna altijd ten koste van het prestatievermogen en maakt structurele verbeteringen onmogelijk. In onze ervaring blijkt steeds opnieuw hoe belangrijk goede organisatorische communicatie³⁵ is voor evidence-based managementinitiatieven.

Een hogere kennisscore voorspelde een daling van het aantal klachten

34. P. Morrel-Samuels, 'Getting the Truth into Workplace Surveys', *Harvard Business Review*, 80/2, februari 2002, p. 111-118.

Gebrek aan steun van het topmanagement

Als het topmanagement nauwelijks steun biedt of zich helemaal niet laat zien, dan zal het de progressieve krachten in de onderneming ontbreken aan de benodigde middelen en bestuurlijke invloed om het verzet te overwinnen van de dinosauriërs, slangen en luiaards die, al dan niet opzettelijk, ieder project zullen saboteren dat matige prestaties blootlegt.

Foutief of inconsistent berekende gemiddelden

Gegevens tot gemiddelden verheffen kan een onverwacht verraderlijk probleem zijn voor een programma voor evidence-based management. In een grote organisatie worden verschillende berekeningsmethoden gebruikt en dat kan tot substantiële vertekeningen leiden. Gemiddelden dienen 'algemene gemiddelden' te zijn, direct afgeleid uit de ruwe basisgegevens, en geen gemiddelden van gemiddelden, een probleem dat pijnlijk acuut kan worden als sommige cellen in een dataset leeg zijn. De oplossing is om consistente rekenprocessen te gebruiken, en labels waarmee expliciet onderscheid wordt gemaakt tussen algemene gemiddelden, en gemiddelden van gemiddelden. Dit is weliswaar geen fatale fout die bij veel organisaties voorkomt, maar we hebben meegemaakt hoe slopend het effect ervan is op een evidence-based besluitvorming, onder meer omdat volhardende en uitgesproken kritiek (hoe ongefundeerd dan ook) over kleine onzorgvuldigheden funest kan zijn voor de geloofwaardigheid van een programma, ook al is het nog zo succesvol.

Dubbelzinnige, verwarrende of vertekende presentatie van uitkomsten

Kleine en onbetekenende beslissingen over de opmaak en weergave van gegevens kunnen van grote invloed zijn op het begrip, het vermeende belang en het onthouden van de betreffende informatie. Er wordt niet voor niets al ruim zeventig jaar empirisch onderzoek gepubliceerd naar het inzicht in en herinneren van kwantitatieve grafieken – zaken die van doorslaggevend belang zijn in een onderneming aangezien beslissingen van het topmanagement vaak zwaar, en soms uitsluitend, steunen op grafieken.

Geen begrip van oorzakelijk bewijs op het betreffende gebied

Zoals het onderzoek naar quasi-experimenten laat zien, zijn er voor experimenten onder strikt gecontroleerde laboratoriumcondities veel goede al-

ternatieven. Bij het interpreteren van bewijs van een oorzakelijk verband uit dergelijk toegepast onderzoek dienen de juiste voorbehouden te worden gemaakt, zodat het bewijs niet wordt overtrokken of meteen wordt weggewuifd. Bovenal dient de lezer erop te worden gewezen dat, ook al heeft het experiment dan wellicht niet onder volledig gecontroleerde omstandigheden plaatsgevonden, er toch veel te leren valt over de relaties tussen variabelen.

Kenmerken van doeltreffende evidence-based managementprogramma's

We hebben in deze korte analyse geprobeerd om een duidelijk inzicht te verschaffen in de instrumenten, methoden, valkuilen, voordelen en veel voorkomende resultaten van evidence-based managementprogramma's met behulp van een samengevoegde dataset. Wij menen dat dergelijke programma's bij uitstek waardevol zijn. Ze behoren namelijk tot de weinige initiatieven in het bedrijfsleven die keer op keer hoog scoren op de volgende aspecten:

- *Alomvattend*: samengevoegde datasets komen van overal uit de hele onderneming – zelfs onderdelen waar het topmanagement doorgaans misschien nauwelijks naar kijkt;
- *Relevant*: samengevoegde datasets weerspiegelen onvermijdelijk de prestaties van allerlei directieleden, managers en individuen die een bijdrage hebben geleverd;
- *Actueel*: samengevoegde datasets blijven actueel omdat ze voortdurend worden aangevuld;
- *Nuttig*: samengevoegde datasets bieden objectieve en praktische beoordelingen van programma's;
- *Winstgevend*: samengevoegde datasets maken het heel lastig om programma's met onverwachte negatieve effecten te verstoppen achter holle retoriek;
- *Informatief*: Samengevoegde datasets leveren objectieve gegevens die onmisbaar zijn voor aanvullende activiteiten, bijvoorbeeld bedrijfsinitiatieven rangschikken naar prioriteit, of de ROI berekenen voor immateriële zaken;
- *Transparant en gemakkelijk te begrijpen*: samengevoegde datasets putten uit gewoon gezond verstand en onderzoek naar de conclusies die niet-statistisch onderlegde mensen trekken als zij kwantitatieve gegevens proberen te begrijpen.
- *Ethisch evenwichtig*: samengevoegde datasets

Evidence-based managementprogramma's met samengevoegde datasets scoren steeds hoog op tal van belangrijke criteria

35. P. Morrel-Samuels en B. Maini, 'Cascading Improvements in Communication Throughout the Workplace', voordracht aan de jaarvergadering van de American Psychological Association, New Orleans, 2006.

bieden een nauwkeurige en rigoureuze analyse van verbanden, zodat tegelijkertijd zowel de winstgevendheid als de bedrijfscultuur kunnen worden verbeterd;

- *Methodologisch degelijk*: met evidence-based management op grond van samengevoegde datasets stijgt het gebruik van ‘best practices’ in toegepast onderzoek, zoals pretest-posttest-vergelijkingen van willekeurig gekozen onderzoeksobjecten, meenemen van passende co-variabelen, en breed systeemdenken.

Slotopmerking

‘Als je datgene waarover je praat, kunt meten en in een getal kunt uitdrukken, dan weet je er iets van; kun je dat niet, dan is je kennis mager en onbevredigend.’ Aldus Lord Kelvin over het belang van

metingen. Toch zou het onzinnig zijn om alles te meten wat in een grote onderneming van belang zou kunnen zijn, in de hoop daar ooit enkele inzichten uit te kunnen winnen. Veel mensen zullen ooit wel eens aan den lijve hebben ervaren hoe vervelend het is om te moeten samenwerken met technocraten die alles meten maar slechts weinig begrijpen. Wij bepleiten hier een verstandig gebruik van samengevoegde datasets en een voortdurende keuze voor evidence-based management. Dan kan met degelijk organisatieonderzoek gelijktijdig de winst en de bedrijfscultuur worden verbeterd. Al diegenen die bij de onderneming betrokken zijn – medewerkers, aandeelhouders, klanten en de samenleving als geheel – hebben daar recht op.

Translated and reprinted by permission of California Management Review. This article was originally published under the English title ‘Merged Datasets: An Analytic Tool for Evidence-Based Management’, by Palmer Morrel-Samuels, Ed Francis and Steve Shucard, in the Fall 2009 issue of the California Management Review, Vol. 52, No. 1, pp. 120-139. Copyright © 2009 by The Regents of the University of California; all rights reserved. Vertaling: Raymond Gijsen.