



De waarde van vergelijkend onderzoek

Auteur(s):

F. Bal en J. Vleugel

De auteurs zijn onafhankelijk wetenschappelijk adviseur respectievelijk senior onderzoeker bij Onderzoeksinstituut OTB, TU Delft

j.vleugel@otb.tudelft.nl

Verschenen in:

ESB, 90e jaargang, nr. 4460, pagina 234, 20 mei 2005

Rubriek:

Vakontw ikkeling

Trefwoord(en):

Als meerdere onderzoeken naar een bepaald onderwerp gedaan zijn, kan het informatief en kostenbesparend zijn deze met elkaar te vergelijken en zo tot een synthese te komen. Dat kan handmatig of met behulp van geavanceerde technieken, zoals meta-analyse en empirisch-bayesiaanse statistiek.

Synthese van studies leidt tot meer inzicht in een empirisch fenomeen. Essentieel hierbij is dat de betreffende studies voldoende transparant zijn, dat wil zeggen dat de veronderstellingen, de analysemethoden, de (eventuele) modellen en de data (direct) toegankelijk, volledig, correct, herleidbaar en toetsbaar moeten zijn voor de lezer. In de praktijk is meestal slechts een deel hiervan transparant. De lezer krijgt een black box voor zich en moet er maar op vertrouwen dat de auteurs goed bezig zijn geweest. Daarbij kunnen de gebruikte onderzoeksmethoden en de aanpak van de (empirische case) studie zodanig verschillen dat de herleidbaarheid van de studies wordt bemoeilijkt (Vleugel, 2000).

Hoe studies vergelijken?

Er bestaan verschillende methoden om studies met elkaar te vergelijken. Vrij algemeen gebruikt is de literaire kwalitatieve methode of literatuurstudie. Daarnaast zijn er kwantitatieve methoden, waartoe met name meta-analyse en bayesiaanse statistiek gerekend kunnen worden. In dit artikel vergelijken wij deze drie methoden op een niet-technische manier. Hierbij staan de volgende vragen centraal:

- » Hoe transparant zijn deze methoden?
- » Is de vergelijking van de studies afhankelijk van de manier waarop ze worden geselecteerd?
- » Is het aantal geselecteerde studies van belang voor de uitkomst van de vergelijking?
- » Is het mogelijk om uitspraken te doen over nog niet onderzochte gebeurtenissen?

Eerst wordt aangegeven wat de methode inhoudt en wat de voor- en nadelen zijn. Vervolgens wordt ingegaan op de vraag wat de meest geschikte toepassingen zijn. Ten behoeve van dat laatste is bij de kwantitatieve methoden ook een aantal praktijkvoorbeelden opgenomen.

Literaire methode

Literatuurstudies zijn studies die de resultaten en consequenties van voorgaande studies bespreken en op elkaar betrekken. Hierbij worden lijsten, tabellen en andere categorische indelingen als ondersteuning van het betoog gebruikt. Van een statistische analyse is geen sprake. Het belangrijkste voordeel van deze methode is de betrekkelijke eenvoud. Er is geen kennis van kwantitatieve methoden en technieken nodig. Dit verklaart ook meteen de toegankelijkheid van de literatuurstudie voor de gemiddelde lezer. De methode kent echter ook belangrijke nadelen; zie Button (1995) voor een uitgebreid overzicht. In de eerste plaats is het selecteren van de studies een subjectief proces, omdat harde selectiecriteria ontbreken. Vaak wordt, vanwege tijdsbeperking en/of financiële restricties, genomen wat direct en vrij voorhanden is (beschikbaarheids- of gemaksbias). De methode heeft vooral een beschrijvende waarde: verklarende analyse van verschillen tussen studieresultaten is erg lastig. Dan is er de 'meerderheidsbias': als maar voldoende studies tot dezelfde conclusie komen, wordt deze conclusie voor waar aangenomen, terwijl niemand zich meer afvraagt wat de kwaliteit van de data, de gemaakte veronderstellingen of de gebruikte onderzoekstechnieken ook al weer was. Niettemin kan er wel een verband bestaan tussen de gebruikte methode en de bereikte resultaten. Deze meerderheidsbias is gerelateerd aan de publicatiebias, waaronder wordt verstaan dat veel auteurs bepaalde resultaten minder snel zullen publiceren als zij afwijken van wat algemeen gangbaar of acceptabel wordt geacht. Ten slotte is er het probleem dat een auteur een grote hoeveelheid studies met verschillende uitkomsten op een kwalitatieve manier op elkaar moet betrekken. Dit is in de praktijk vaak te veel gevraagd.

Meta-analyse

Meta-analyse stamt uit de geneeskunde. Medici willen weten welke behandelmethoden effect hadden of (zullen) hebben, in welke mate,

of er contra-indicaties zijn, etc. Datasets zijn in het algemeen eenvoudiger te verkrijgen dan in de economische wetenschap, omdat data vaak routinematig worden verzameld volgens strikte standaardprocedures en vaak ook over lange perioden beschikbaar zijn. Medisch statistici gebruiken vervolgens licht aangepaste klassieke statistische technieken om genoemde medische vragen te beantwoorden (Jones, 1995).

De methode werkt als volgt. Er worden zoveel mogelijk studies verzameld, waaruit de bijbehorende waarden van de variabelen worden gedestilleerd. Deze data worden in een rekenprogramma ingevoerd, dat per variabele tabellen maakt over alle studies. Vervolgens kunnen op deze uitkomsten statistische bewerkingen worden losgelaten, van eenvoudig (gemiddelden en spreiding) tot geavanceerd (correlaties bepalen etc.).

Praktijkvoorbeelden

Teneinde meta-analyse en bayesiaanse statistiek intrinsiek met elkaar te kunnen vergelijken, zijn natuurwaarderingsstudies zeer aantrekkelijk, omdat aan andere methoden belangrijke bezwaren kleven. Bij gebrek aan een markt voor 'milieugoederen' kunnen 'milieuprijzen' niet rechtstreeks worden bepaald. Een alternatieve aanpak is nodig (zie ook Vleugel & Bos, 2004). Dit kan op basis van een compleet nieuwe studie of bestaande studies. Het eerste kost vrij veel tijd en geld en is alleen mogelijk voor reeds bestaande natuurgebieden. Het tweede komt neer op het 'vertalen' van de resultaten van studies voor andere natuurgebieden naar het gebied dat gewaardeerd moet worden (Loomis, 1992). Vergelijkende milieu-economische studies zijn in de VS verplicht bij het economisch waarderen van natuurgebieden. Dit soort studies is ook in Nederland uitgevoerd. Hierbij is met behulp van stated preference studies (mondelinge of schriftelijke enquêtes) de betalingsbereidheid van bezoekers (uitgedrukt in gemiddelde (average) willingness to pay; aWTP-waarde) van bekende natuurgebieden bepaald als benadering van de werkelijke waarde van deze natuurgebieden. Omdat natuurgebieden vaak niet identiek zijn, wordt getracht de verschillen in gebiedskarakteristieken te vertalen in een gecorrigeerde aWTP-waarde voor het studiegebied.

De aWTP-waarde kan ook via een empirisch-bayesiaanse studie worden bepaald. Hierbij wordt een natuurwaarderingsfunctie geschat. De variabelen in deze waarderingsfunctie zijn dezelfde als die in andere waarderingsstudies zijn gebruikt. Om de waarden van de variabelen te bepalen, heeft Bal een database van 33 eerder uitgevoerde waarderingsstudies opgebouwd (Bal, 2002). De betreffende gegevens zijn ingevoerd in softwarepakket S-Plus. Dit pakket heeft de posteriors bepaald door de 33 studies met elkaar te vergelijken. De gebruikte datasets, toegepaste methodieken, bestudeerde tijdsperioden en natuurgebiedskarakteristieken leverden een bonte verzameling aan bruikbare en minder bruikbare gegevens op.

Twee variabelen bleken significant, namelijk het gemiddeld inkomen van (potentiële) bezoekers en de omvang van het natuurgebied (zie [tabel 1](#)).

Tabel 1. Geschatte waarden op basis van de dataset

	parameter- waarde	standaard- fout	t-waarde	eenzijdige t-toets
intercept (0)	-153,5222	79,00560	-1,9432	2
gemiddeld inkomen (β_0)	0,0080	0,00312	2,5773	1
omvang (β_0)	0,0002	0,00004	4,2358	1

Uit kolom 1 van deze tabel kan de volgende waardefunctie worden afgeleid: aWTP_i is gelijk aan de statistisch significante (de constante) -153,52 plus 0,008 maal het gemiddeld inkomen plus 0,0002 maal de omvang van het gebied *i*. Het beperkte aantal bruikbare (variëaties in) casestudies verhindert momenteel dat meer variabelen gedestilleerd kunnen worden, die kunnen bijdragen aan een verfijndere aWTP_i-schatting (Bal, *ibid*).

Met deze geschatte waarderingsfunctie is het nu mogelijk om

a. de gecorrigeerde (posterior) casestudieresultaten van elk van de 33 studies te bepalen;

b. de waarde van een nog niet eerder gewaardeerd natuurgebied te schatten aan de hand van de (geplande) karakteristieken van dit gebied.

Ad a. De resultaten van elk van de 33 case-studies in de dataset zijn te vergelijken met de waarde uit de waarderingsfunctie. Dan blijkt bij de:

» Oostvaardersplassen de synthetisch geschatte aWTP-waarde 16% lager te liggen dan die uit de casestudie (Groot & van Buiren, 1998).

» Waddenzee de synthetisch geschatte aWTP resp. 61% en 98% hoger is dan de aWTP-schatting uit de casestudies (AIDEnvironment, 1999; Spaninks et al., 1996).

De studie naar de Oostvaardersplassen bevat dus een overschatting en de studies naar de Waddenzee onderschatten de

waarde van deze natuurgebieden volgens bezoekers.

Ad b. In Bal (ibid) is een toepassing opgenomen, een plan voor de aanleg van het Kuindermeer en daarvan de groene variant met ongeveer 3300 hectare nieuwe natuur (Bos & van Leeuwen, 2001). Bij een gemiddeld inkomen van bezoekers van € 20130 per jaar geeft de synthetisch geschatte functie een aWTP-waarde voor het Kuindermeer van € 8,18 per bezoekend Nederlands huishouden per jaar.

De methode heeft diverse voordelen. In de eerste plaats maakt het gebruik van getallen kwantitatieve verklarende analyse mogelijk. Hierbij kan de kwaliteit van de data en onderzoekstechnieken onderzocht worden. De inzet van een computer vermijdt de cognitieve beperkingen van de menselijke geest. Deze methode van synthetiseren is dus effectiever en efficiënter. Een ander voordeel is dat de methode gebruikt kan worden om voorspellingen te genereren en daarbij inzicht te bieden in fenomenen waarvoor geen specifieke studies bestaan (Button, 1995).

De methode kent ook beperkingen en nadelen. In de eerste plaats wordt een zekere mate van standaardisatie in rapportage verondersteld. In de geneeskunde is dit gebruikelijker dan in de economie. In de tweede plaats is ook bij deze methode het verzamelen van data en destilleren van informatie daaruit nog steeds een subjectief proces. Gebruik van statistische methoden vermindert echter de mate van subjectiviteit en vergroot de robuustheid van de conclusies.

Empirisch-bayesiaanse statistiek

Wetenschappelijke kennis groeit met de tijd. Nieuwe data leiden tot nieuwe hypothesen, herformulering van onderzoeksvragen en uiteindelijk aangepaste conclusies. Bij het beoordelen van de huidige situatie alsmede ten behoeve van het opstellen van voorspellingen over gebeurtenissen die nog plaats moeten vinden, kan bayesiaanse statistiek met succes ingezet worden. De methode biedt een logisch, kwantitatief raamwerk voor evaluatie van oude en nieuwe data en wordt op veel terreinen toegepast. Het is in feite een speciale vorm van waarschijnlijkheidsberekening, die niet werkt volgens de veel bekendere klassieke statistiek die gebaseerd is op frequentieverdelingen (Bayesianorg, 2005).

Methoden

De onderzoeker gaat als volgt te werk. Hij verzamelt zoveel mogelijk bestaande studies, destilleert daaruit alle beschikbare data alsmede meta-informatie over de betreffende studie en voert deze in in een softwarepakket. Vervolgens bepaalt het pakket de meest waarschijnlijke waarden van een verzameling modelvariabelen. Ten opzichte van de klassieke statistiek zijn hierbij twee zaken uniek. In de eerste plaats gaat bayesiaanse statistiek ervan uit dat elke gebeurtenis uniek is. In de tweede plaats spelen de verwachtingen van de onderzoeker een belangrijke rol, met name in het begin van de analyse. De onderzoeker voert namelijk zelf voor elke variabele een startwaarde (interval) in, de zogenoemde prior. De software bepaalt vervolgens de meest waarschijnlijke waarde voor elke variabele, de zogenoemde posterior, de gezochte statistische waarde.

Het bepalen van de prior vereist voldoende kennis van statistische methoden en technieken en van de toepassing. De keuze van de beginwaarde is een persoonlijke keuze, echter deze heeft geen invloed op het verloop van het proces als er maar voldoende studies beschikbaar zijn. De theoretisch optimale waarde wordt dan zeer dicht benaderd. Bayesiaanse statistiek kent een empirische variant - de empirisch-bayesiaanse statistiek - waarbij de waarde van priors wordt bepaald zonder interventie van de onderzoeker.

Evaluatie

De empirisch-bayesiaanse statistiek kent meerdere voordelen vergeleken met de klassieke statistiek. Bij deze methode zijn veel minder studies nodig om een optimale oplossing te bepalen, omdat de methode niet alleen inhoudelijke informatie (data), maar ook meta-data, zoals informatie over de auteur(s) gebruikt. In de tweede plaats kan later aanvullende, externe informatie (uit nieuwe studies) toegevoegd worden. Hiermee wordt de nauwkeurigheid van de schatter (posterior) vergroot. Dan is er de mogelijkheid om theorieën te toetsen en te vergelijken. Ook kan de methode worden gebruikt om de toepasbaarheid van rekenmodellen te testen. Verder is te bepalen hoeveel studies nodig zijn voor een optimale uitkomst gegeven de kosten per studie. Zo kan dus worden bepaald of het aan de dataset toevoegen van een extra studie zinvol is. Ook kunnen volstrekt verschillende datasets worden gepoold om dwarsverbanden boven water te krijgen. Ten slotte is ook een sterk punt dat met de methode op relatief eenvoudige wijze grote en complexe databases met zeer uiteenlopende studies gesynthetiseerd kunnen worden, wat met de klassieke statistiek niet of nauwelijks mogelijk is.

Een nadeel van de methode is dat het (zeker bij grote datasets) onbegonnen werk is om de bijbehorende berekeningen zonder moderne software uit te voeren. Dit verklaart waarom de methode in de negentiende eeuw naar de achtergrond is geraakt en pas sinds de opkomst van de pc begin jaren tachtig weer sterk in opkomst is.

Toepassingen binnen de economie

Datasets zijn vaak een sluitpost in economisch onderzoek. Dit leidt er mede toe dat data ad hoc worden verzameld, datasets incompleet zijn en inconsistenties bevatten. Verder ontbreekt standaardisatie bij de toepassing van methoden, de naamgeving van variabelen en hun definitie en verslaggeving. Dit zijn problemen die vooral kwantitatieve analyse parten spelen.

Hoe wordt daar in de praktijk mee omgegaan? Als we kijken naar toepassingen van meta-analyse, dan zijn deze vooral binnen de ruimtelijke en de milieueconomie te vinden (De Groot et al., 2002). Het idee dat meta-analyse de bovengenoemde problemen met data relatief eenvoudig kan oplossen, blijkt helaas niet waar te zijn, want meta-analyse kan (Vanhonacker et al., 1992):

» geen gaten in een dataset opvullen, noch de kwaliteit van de data verbeteren;

» niet goed omgaan met significante verschillen in de karakteristieken van de casestudies (diversiteit). Vaak is het daarom niet mogelijk

om studies zinvol te vergelijken;

» casestudies met onvolledige beschrijvingen niet upgraden. Als er een black box is, dan blijft deze bestaan waardoor synthetiseren niet mogelijk is.

Er kunnen vraagtekens gezet worden bij de selectie van studies en de manier waarop de data in ons bekende, openbare toepassingen in Nederland zijn geanalyseerd. Zo leidt het gebruik van gemiddelden tot het 'wegmoffelen' van de verschillen in karakteristieken van individuele studies. In een synthese worden deze verschillen nu juist gebruikt om verschillen in studieresultaten te verklaren. Het onder deze condities vergelijken van eerder ondernomen casestudies heeft weinig zin (Bal, 2002).

Conclusie

Het vergelijken van studies teneinde de meest waarschijnlijke resultaten te kunnen bepalen is een noodzakelijk, maar ingewikkeld proces. De tot nu toe gebruikte vergelijkende literatuurstudies bevatten (te) veel valkuilen, terwijl bij kwantitatieve vergelijkende meta-analytische studies de grote variatie in de studiekarakteristieken de succesvolle toepassing van deze methodiek (te) vaak in de weg staat. Meta-analyse, gebaseerd op klassieke statistiek, kan tevens weinig uitrichten bij kleine aantallen beschikbare studies en/of een verzameling studies met een laag empirisch gehalte; een veel voorkomend probleem in het toegepast economisch onderzoek. Bayesiaanse statistiek kan hiermee wel omgaan. Op basis van een relatief kleine set studieresultaten kan bijvoorbeeld een functie geschat worden waarmee een natuurgebied kan worden gewaardeerd. De resultaten zijn completer en robuuster dan bij meta-analyse, terwijl de kosten van een bayesiaanse studie relatief beperkt en vooraf calculeerbaar zijn. De methode is ook toepasbaar voor andere, inclusief nog aan te leggen, natuurgebieden.

Frans Bal en Jaap Vleugel

Literatuur

AIDEnvironment (1999) *De schaduwkant van het Waddengas*. Amsterdam.

Bal, F. (2002) *Valuing Dutch nature areas: A different approach*. Maastricht: Shaker Publishing.

Bayesianorg (2005) *What is Bayesian analysis?* www.bayesianorg/bayesexp/bayesexp.htm

Bos, E.J. & M.G.A. van Leeuwen (2001) *Ontwikkeling van een instrumentarium voor het bepalen van regionaal-economische effecten van landinrichtingsprojecten; Illustratie aan de hand van de case 'het Kuindermeer'*. Den Haag: LEI.

Button, K.J. (1995) What can Meta-analysis Tell us about the Implications of Transport? In: *Regional Studies*, 29, 1995, 507-517.

Groot, A.W.M. & K.H.S. van Buiren (1998) *Natuurlijk vermogen: een empirische studie naar de economische waardering van natuurgebieden in het algemeen en de Oostvaardersplassen in het bijzonder*. SEO-rapport, 475, Amsterdam: Universiteit van Amsterdam.

Groot, H.L.F. de, R.J.G.M. Florax & R.A. de Mooij (2002) Meta-analyse: Analyseer de analyses. *ESB*, 1 februari, 96-98, Rotterdam.

Jones, D.R. (1995) Meta-analysis weighing the evidence. In: *Statistics in Medicine*, 14, 137-149, Wiley, Chichester.

Loomis, J. B. (1992) The evolution of a more rigorous approach to benefit transfer: benefit function transfer. *Water Resource Research*, 28, 701-705, Washington D.C: AGU Publications.

Spaninks, F.A., O.J. Kuik, & J.G.M. Hoogeveen (1996) *Willingness to pay of Dutch households for a natural Wadden Sea: An application of the Contingent Valuation method*. Mimeo E96/6, Amsterdam: Institute for Environmental Studies.

Vanhonacker, W.R., & L.J. Price (1992) Using meta-analysis results in a Bayesian updating: The empty cell problem. *Journal of Business and Economic Statistics*, 10, IV, 427-435.

Vleugel, J.M. (2000) *Design of transport and land-use scenarios*. Principles and applications. Dissertatie, Amsterdam: Vrije Universiteit.

Vleugel, J., & E. Bos (2004) Rondje Randstad is slecht voor de natuur. *ESB*, 20 augustus, 392-394.