

1 antidiscriminatiewet). Zie: P. Taelman, "De stakingsvordering in de antidiscriminatiewet. Een maat voor niets?", *NjW*, nr. 69, 2004, 511-512.

<sup>19</sup> Zie: E. Brems, *NjW*, nr. 60, 2004, 206.

<sup>20</sup> Voorz. Kort Ged. Gent 31 december 2003, *NjW*, nr. 60, 2004, 205-206.

<sup>21</sup> Zie kritisch: D. De Prins, "Gediscrimineerd homopaar krijgt geen rechtsherstel", *De Juristenkrant*, 2004, nr. 84, 1 & 12.

<sup>22</sup> D. Minten, "Veel discriminatieklachten missen grond", *De Standaard*, 26 februari 2004. Zie m.b.t. de 'zaak Joos' o.m.: S. Sottiaux & J. Vrielink, "Acht redenen waarom Joos niet veroordeeld mag worden", *De Juristenkrant*, nr. 83, 2004, 2-3; M. Elcharidus, "Het waarde- en normendebat", *Knack*, 19 februari 2004.

<sup>23</sup> Instituut voor Vreemdelingenrecht en Rechtsantropologie, Katholieke Universiteit Leuven.

---

## CRIMINOLOGIE EN METHODOLOGIE

---

### DE VANGST-HERVANGST METHODE VOOR DE SCHATTING VAN DELINQUENTE POPULATIES

#### I. Introductie

Criminologie wordt als wetenschap of studie van de criminaliteit al in zijn definitie geconfronteerd met wat allicht het kernprobleem van de discipline genoemd kan worden. Met name het inschatten van de omvang van de criminaliteit, van wat als deviant of delinquent gedrag gecatalogeerd staat, kan worden aanzien als een van de, zonet het centrale pijnpunt van de discipline, en dit sedert het ontstaan ervan tot op de dag van vandaag. Veelal wordt hiernaar verwezen als de 'dark number' problematiek. Victim-survey's en self-report studies zijn twee gekende en gewaardeerde manieren om hieraan, uiteraard slechts gedeeltelijk, tegemoet te komen. Bij een victimsurvey of slachtofferenquête poogt men tot een meer accurate schatting te komen van die criminaliteit die wel een slachtoffer heeft, maar geen aangifte en dus ook geen officiële registratie kent. Self-report studies zijn dan weer in staat een licht te werpen op de daders van criminaliteit die niet noodzakelijk een slachtoffer hebben. Maar beide technieken hebben hun eigen specifieke nadelen, al was het

maar omdat men in belangrijke mate afhankelijk is van de bereidwillige medewerking van respondenten die, hetzij als slachtoffer, hetzij als dader, zeker niet steeds geneigd zijn of er baat bij hebben het 'volledige' verhaal te doen. Onder meer om deze reden, en tal van andere argumenten die genoegzaam gekend zijn (Van Kerckvoorde, 1995), kunnen beide methoden steeds slechts een partiele opheldering van het 'dark number' realiseren.

Deze bijdrage heeft als doel in een inleiding en illustratie te voorzien van een alternatieve methodiek om, complementair aan en naast de bovenvermelde gekende technieken, tot een betere schatting van delinquente populaties te komen. Het gaat met name om een aan de biologie ontleende techniek, de vangst-hervangst methode, toegepast binnen de criminologie; een schattingsprocedure voor delinquente of illegale populaties waarbij de officieel geregistreerde criminaliteit of politiestatistiek het uitgangspunt vormen. In wat volgt staan we kort stil bij de essentie van de vangst-hervangst methode aan de hand van een illustratie uit de discipline waar deze schatting ontstond, de biologie; daarna wordt deze illustratie vertaald naar en geëvalueerd op zijn bruikbaarheid binnen de criminologie, o.a. aan de hand van een op dit moment lopend onderzoek in Nederland<sup>1</sup>, om tenslotte het geheel opnieuw te kaderen binnen het hierboven geschetste verhaal.

#### II. Vangst-hervangst methode

De vangst-hervangst methode voor het schatten van populaties is afkomstig uit de biologie waar het gebruikt wordt om een beeld te krijgen van de populatieomvang van een bepaalde diersoort binnen een welomschreven gebied. Een eenvoudige illustratie stelt ons allicht best in staat zowel de idee als de schattingsprocedure hierachter duidelijk te maken. Voor andere voorbeelden of een bredere uiteenzetting van wat hier bij wijze van introductie summier geïllustreerd wordt, verwijzen we o.a. naar Cruijff & van der Heijden (2003a), Cruijff & van der Heijden (2003b), Cruijff e.a. (2002). Het is met name ook op deze publicaties dat we ons baseren voor onderstaande uiteenzetting.

Veronderstel, men is geïnteresseerd in het precieze aantal van een bepaalde vissoort dat een vijver telt. Op  $n$  verschillende tijdstippen in een bepaald periode gaat men vangen, wordt de vangst gemarkeerd, en gooit men de vissen terug. Het resultaat van deze procedure ziet er dan als volgt uit:

aantal keer gevangen	1	2	3	...	...	$n$
aantal vissen	$f(1)$	$f(2)$	$f(3)$	...	...	$f(n)$ (1)

$f(\text{niet gevangen}) = f(0)$   
 $f(\text{gevangen}) = f(1) + f(2) + f(3) + \dots + f(n)$  (1)  
 $f(\text{totaal}) = f(\text{niet gevangen}) + f(\text{gevangen})$  (2)  
 Het totale aantal vissen in de vijver kan dus worden bekomen door de niet geobserveerde vissen, zijnde  $f(\text{niet gevangen})$  of  $f(0)$ , te sommeren met de geobserveerde of gevangen vissen, zijnde  $f(\text{gevangen})$ . Het tweede deel van deze sommatie,  $f(\text{gevangen})$ , zelf de som van  $f(1), f(2), \dots, f(n)$ , is gekend en vormt geen probleem. Het eerste deel,  $f(\text{niet gevangen})$  is echter een onbekende, maar kan worden geschat onder de aanname dat bovenstaande cijferreeks (1) een Poisson verdeling volgt. Bij de assumpties die horen bij een dergelijke verdeling staan we straks nog even stil.

De Poisson verdeling, verwant met de klassieke binomiaalverdeling, baseert zich op de stochastische veranderlijke die het aantal keren 'succes' ( $p$ ) telt bij  $n$  onafhankelijk herhaalde Bernoulli experimenten, en dit specifiek voor situaties waarbij  $n$  groot is en  $p$  klein (Maes e.a., 2000). Dit,  $n$  groot en  $p$  klein, is zo voor onze 'vissen in de vijver' illustratie, maar gaat eveneens op voor de meeste toepassingen van de vangst-hervangst methode binnen de criminologie.

In de veronderstelling dat over onze vangsten heen de kans om één bepaalde vis te vangen constant is, dan verdeelt de kans om die vis tijdens elke vangst opnieuw te vangen, zich als een Poisson variabele met deze kansverdeling:

$$p(y; m) = \frac{e^{-m} m^y}{y!} \quad (3)$$

Hierbij is  $m$  de Poisson parameter<sup>2</sup> en  $y$  het aantal keer dat men die bepaalde vis gevangen heeft. In de veronderstelling dat er in een

bepaald jaar 6 'vangsten' zijn, dan vinden we volgende waarden:

aantal keer gevangen	0	1	2	3	4	5	6
kans (met $m=.75$ )	.472	.354	.133	.033	.006	.001	.000
kans (met $m=1$ )	.368	.368	.184	.061	.015	.003	.001

Indien de Poisson parameter  $.75$  is, dan is de kans dat een bepaalde vis éénmaal gevangen wordt  $.354$ , tweemaal  $.133$ , driemaal  $.033$ ,... Diezelfde verdeling bij een hogere Poisson parameter ( $m=1$ ), verhoogt de relatieve waarschijnlijkheid om een bepaalde vis meerdere malen te vangen, al blijft die kans in absolute waarden nog steeds klein. In de praktijk is de Poisson parameter natuurlijk niet gegeven; het is echter mogelijk een parameter te schatten die zo goed mogelijk aansluit bij de frequenties of cijferreeks die we wel kennen, met name  $f(1), f(2), \dots$  (zie (1)). De geschatte Poisson parameter kan vervolgens in formule (3) worden ingevoerd en op die manier zijn we ook in staat  $f(0)$  of  $f(\text{niet gevangen})$  te berekenen. Deze frequentie integreren in (2) levert ons tenslotte een schatting op van het totale aantal vissen in de vijver, de beoogde doelstelling

### III. Toepassingsgebied binnen de criminologie

Hierboven haalden we al even aan dat aan de veronderstelling dat een bepaalde cijferreeks een Poisson verdeling volgt, enkele assumpties verbonden zijn. Deze assumpties worden hier even van naderbij bekeken, juist omdat met name hier de analogie tussen bovenstaande illustratie en het toepassen van de vangst-hervangst methode voor criminologisch relevante vraagstukken, onder druk komt te staan. Hiervoor inspireren we ons op bestaand afgerond en lopend criminologisch onderzoek in Nederland waarbij gebruik wordt gemaakt van de vangst-hervangst methode; het gaat hierbij om de schatting van het aantal illegale vuurwapens in Nederland, om het aantal illegale vreemdelingen, en om het aantal bestuurders in dronken toestand (Cruijff & van der Heijden, 2003a; 2003b; Cruijff e.a., 2002; Spapens & Bruinsma, 2002).

Een eerste assumptie werd hierboven reeds impliciet verondersteld; de Poisson parameter is

voor alle eenheden identiek, wat inhoudt dat alle eenheden theoretisch gezien éénzelfde pakkans hebben. Men spreekt ook over de assumptie van een homogene populatie. Alhoewel dit voor vissen in een vijver misschien nog aannemelijk is, voor toepassingen binnen de criminologie is het toch bedenkelijk. Cruijff e.a. (2002) verwijzen in hun onderzoek naar het schatten van het aantal illegale vreemdelingen in Nederland, dat sommige illegalen nu eenmaal zichtbaarder zijn dan anderen, en daardoor een grotere kans lopen op 'gepakt' te worden. Maar natuurlijk kan deze pakkans ook verschillen van stad tot stad, of is ze afhankelijk van de prioriteiten die de politie zich stelt, en die dus van korps tot korps kunnen variëren. Voor het probleem van een heterogene populatie is er evenwel een oplossing; het is mogelijk de Poisson parameter als een functie van een aantal verklarende variabelen, covariaten, te specificeren. Dit noemt men Poisson regressie<sup>3</sup>. Op die manier wordt de heterogeniteit aan 'pakkansen' afhankelijk gesteld van enkele factoren, en geïncorporeerd in het model. Bij het schatten van de populatie illegalen wordt dan bijvoorbeeld gedacht aan 'de reden van staandehouding', 'de leeftijd', 'het geslacht' en 'het land van herkomst' (Cruijff e.a., 2002:51). Natuurlijk blijft de assumptie gelden in die zin dat Poisson parameters voor individuen met identieke waarden op de opgenomen covariaten, homogeen zijn. Men spreekt dan van geobserveerde heterogeniteit. Variabiliteit in de Poisson parameters veroorzaakt door niet in de regressie opgenomen invloeden, ongeobserveerde heterogeniteit blijft problematisch (van der Heijden e.a., 2003)<sup>4</sup>.

Een tweede assumptie stelt dat de Poisson parameter niet mag veranderen over de tijd. Opnieuw, ook hier is de analogie bij toepassingen in de criminologische sfeer niet evident. Bijvoorbeeld in het onderzoek naar illegale vuurwapens is het niet ondenkbeeldig dat iemand die éénmaal werd gepakt met een illegaal vuurwapen voorzichtiger wordt, en het vuurwapen bijvoorbeeld niet meer bij zich draagt, of genoeg onvindbaar opbergt. Maar evengoed is het mogelijk dat de politie zij bij wie ze in het verleden illegale vuurwapens aantroffen, beter in de gaten houden, wat de 'pakkans' verhoogt. Het

eerste geval staat bekend als 'negatieve besmetting', het tweede als 'positieve besmetting' (van der Heijden e.a., 2003). In wezen hoeft de 'pakkans' niet zozeer constant te zijn; de kern van de zaak is dat een verandering in de pakkans niet mag veroorzaakt worden door een eerdere vangst zoals in bovenstaande voorbeelden, omdat zo de eerder vermelde onafhankelijkheid van de herhaalde Bernoulli experimenten in het gedrang komt (Maes e.a., 2000).

Een derde assumptie is hier nauw mee verbonden, en stelt dat de populatieomvang ( $f(\text{totaal})$ ) constant of 'gesloten' dient te zijn. Opnieuw is deze assumptie voor 'vissen in een vijver' niet zo problematisch; bij illegale of delinquente populaties ligt dit echter anders. Illegale vuurwapens worden in beslag genomen, en mogelijk houdt de bezitter er zich in de toekomst niet meer mee bezig. De kans op een hervangst is met andere woorden onbestaande geworden. Maar bijvoorbeeld ook bij een schietincident, waarbij de bezitter van het illegale wapen veroordeeld en gedetineerd wordt en zo ook, toch voor een bepaalde tijd, uit de doelpopulatie 'verdwijnt', is hervangst niet meer mogelijk. Een ander voorbeeld; illegale vreemdelingen worden in sommige gevallen het land uitgezet en komen niet noodzakelijk terug. Of hij die dronken achter het stuur 'gepakt' wordt, kan bijvoorbeeld na het veroorzaken van een ernstig ongeval, een aanzienlijke tijd zijn rijbewijs verliezen of ziet zijn wagen in beslag genomen<sup>5</sup>.

Samenvattend kan dus gesteld worden dat het toepassen van de vangst-hervangst methode voor criminologisch relevante schattingen niet zonder problemen is. Met name geobserveerde heterogeniteit en besmetting vormen bij het schatten van illegale of delinquente populaties een potentiële bedreiging voor de assumpties van de Poisson verdeling. We zijn hier binnen het bestek van deze bijdrage echter noodzakelijk beknopt gebleven; voor een verdere bespreking en volledig uitgewerkte voorbeelden, maar ook voor de constructie van betrouwbaarheidsintervallen rond de populatieschatting, en de fit van het Poisson regressiemodel, waarvoor de ruimte hier ontbrak, verwijzen we de geïnteresseerde lezer naar de reeds vermelde literatuur.

#### IV. Slot

De doelstelling van deze korte bijdrage was de lezer te introduceren in het potentieel van het populatieschatten aan de hand van de vangst-hervangst methode. In het bijzonder ging onze aandacht uit naar mogelijke en relevante criminologische toepassingen, met name naar het schatten van delinquente of illegale populaties. Maar aan elke methode zijn nadelen verbonden, zo ook aan deze. We hadden het met name over de voornaamste bedreigingen voor de centrale assumpties verbonden aan de techniek. Hierbij gaan we allicht voorbij aan wat misschien wel het grootste probleem is bij criminologisch relevante toepassingen, en de analyse voorafgaat; de vangst-hervangst methode vertrekt namelijk van officieel geregistreerde criminaliteitscijfers of politiestatistiek die allesbehalve zijn opgesteld en verzameld met het oog op dergelijke bewerkingen. Dat het gebruiksklaar maken van dergelijke bestanden een tijdrovende en delicate bezigheid is, bleek reeds uit de eigen ervaring bij de preparatie van een geschikt databestand voor het schatten van het aantal illegale vuurwapens in Nederland. Er zijn allicht weinig redenen die doen vermoeden dat dit in België anders zou zijn.

Niettemin is het onze overtuiging dat de vangst-hervangst methode, in combinatie met andere en meer gekende technieken, in een beter zicht op de 'dark number' problematiek resulteert. Hiermee leunen we aan bij een 'milde' visie op triangulatie van methoden of convergente validiteit, "in favour of a complementarity between methods where different kinds of methods reveal different parts and aspects of the same content area" (Waage, 1997; zie ook Hox, 1986).

#### V. Bibliografie

Cruyff, M.J.L.F. & P.G.M. van der Heijden (2003a). *Het schatten van delinquente populaties met de vangst-hervangstmethode*. Paper gepresenteerd op het NVK congres 2003, Amsterdam.

Cruyff, M.J.L.F. & P.G.M. van der Heijden (2003b). *Benutting HKS en ontwikkeling vangst-hervangstmethode. Een onderzoek in opdracht van het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) van het Ministerie van Justi-*

*tie*. Utrecht, Universiteit Utrecht, Onderzoeksschool IOPS afdeling Utrecht (zie ook: [www.wodc.nl](http://www.wodc.nl)).

Cruyff, M., P. van der Heijden & J. van der Leun (2002). Een schatting van het aantal illegale vreemdelingen in Nederland. In: Engbersen, G. R. Staring, J. van der Leun, J. de Boom, P. van der Heijden & M. Cruyff (eds.). *Illegale vreemdelingen in Nederland. Omvang, overkomst, verblijven uitzetting*. Erasmus Universiteit, RISBO (In opdracht van het Ministerie van Justitie).

Hox, J.J.C.M. (1986). *Het gebruik van hulptheorieën bij operationalisering. Een studie rond het begrip subjectief welbevinden*. Academisch proefschrift, Universiteit van Amsterdam.

Maes, C., G. Loosveldt & J. Welkenhuysen-Gybels (2000). *Basisconcepten van de kanstheorie en de inductieve statistiek*. Acco, Leuven.

Spapens, A.C. & M.Y. Bruinsma (2002). *Vuurwapens gezocht. Vuurwapengebruik, -bezit en handel in Nederland, 1998-2000*. IVA Tilburg, Instituut voor sociaal-wetenschappelijk beleidsonderzoek en advies.

Van der Heijden, P.G.M., M. Cruyff & H. van Houwelingen. (2003) Estimating the size of a criminal population from police registrations using the truncated Poisson regression model. *Statistica Neerlandica*, vol. 57, 3, 289-304.

Van Kerckvoorde, J. (1995). *Een maat voor het kwaad?* Universitaire Pers Leuven, Reeks Samenleving, Criminaliteit & Strafrechtspleging, 8, Leuven.

Waage, H. (1997). Is current validation valid? Some intriguing problems with traditional validation designs for general (value-) surveys. *Bulletin de Méthodologie Sociologique*, March 1997, 54, 41-63.

STEFAN BOGAERTS<sup>6</sup>

STEEFAAN PLEYSIER<sup>7</sup>

STIJN VANHEULE<sup>8</sup>

#### Noten

<sup>1</sup> 'Aard en omvang vuurwapencriminaliteit in Nederland 2001-2003', uitgevoerd door het IVA Tilburg in opdracht van het WODC (Ministerie van Justitie, Nederland).

<sup>2</sup> De Poisson parameter wordt geschat met behulp van speciale computerprogramma's;

voor onze illustratie nemen we de cijferreeksen van Cruijff e.a. (2002) over, die uitgedrukt worden als kansen die optellen tot 1.

<sup>3</sup> Poisson regressie, een eenvoudige uitbreiding van de Poisson verdeling in formule (3), ziet er als volgt uit:

$$p(y(i); m(i)) = \frac{e^{-m(i)} m^{y(i)}}{y(i)!}$$

waarbij  $\log m(i) = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_k x_{ki}$  (4)

Elk individu heeft dus een eigen Poisson parameter  $m(i)$ , en deze parameter is voor elk individu dus een functie van een aantal verklarende variabelen  $x(i)$ . Doordat deze verklarende variabelen voor elk individu van invloed zijn op de Poisson parameter, bepalen ze ook de kans dat een bepaald individu 'gepakt' of 'herpakt' wordt ( $y(i)$ ). Ook hier worden de Poisson parameters uiteindelijk ook  $f(0)$  opnieuw geschat op basis van een computerprogramma voor afgeknotte Poisson regressie (Cruijff e.a., 2002).

<sup>4</sup> Echter, ook voor ongeobserveerde heterogeniteit, wat kan leiden tot 'overdispersion', worden oplossingen geformuleerd: "For the truncated Poisson regression model a Lagrange multiplier test on overdispersion was developed by Gurmu (1991). It compares the model fit of the Poisson model against alternative models with an extra dispersion parameter included, such as the negative binomial model" (van der Heijden e.a., 2003:297).

<sup>5</sup> Ook hiervoor biedt Poisson regressie een 'oplossing' (zie formule (4)). De uitval, tijdelijk of definitief, van individuen maakt de  $y(i)$  voor die individuen onbetrouwbaar; de  $x(i)$  waarden zijn dit echter wel. Nu is het mogelijk om voor die individuen waarvan we de  $y(i)$  waarden niet betrouwen, Poisson parameters te berekenen enkel op basis van zij waarvoor we  $x(i)$  en  $y(i)$  wel betrouwen. Die specifieke Poisson parameters worden tenslotte gebruikt voor het berekenen van de kansen. Cruijff e.a. waarschuwen echter dat "niettemin een dergelijke schatting altijd met enige voorzichtigheid moet worden beschouwd" (2002:53).

<sup>6</sup> Stefan Bogaerts is seniorhoofdonderzoeker bij IVA Tilburg, Universiteit van Tilburg en bijz. hoofddocent aan de Faculteit Psychologie, Universiteit Gent. Hij is onder meer promotor van

het onderzoek 'Aard en omvang vuurwapencriminaliteit in Nederland 2001-2003', uitgevoerd door het IVA in opdracht van het WODC (Ministerie van Justitie, Nederland), waarvan sprake in deze bijdrage.

<sup>7</sup> Stefaan Pleysier is wetenschappelijk onderzoeker (FWO) aan de Afdeling Strafrecht, strafverordering en criminologie, Faculteit Rechtsgeleerdheid, van de Katholieke Universiteit Leuven. Hij zetelt als methodoloog in de WODC begeleidingscommissie van het bovenvermelde onderzoek 'Vuurwapencriminaliteit'.

<sup>8</sup> Stijn Vanheule is doctorassistent aan de Faculteit Psychologie, Universiteit Gent.

## VALKUILEN BIJ DE ANALYSE VAN PARKETDATA

In 2003 voerde de onderzoeksgroep *Sociale Veiligheidsanalyse* (UGent - VUB) een studie uit naar de exploitatiemogelijkheden van parketdatatabanken<sup>1</sup>. Tijdens het analyseren van de parketdata werden wij, de onderzoekers<sup>2</sup>, geconfronteerd met enkele belangrijke methodologische valkuilen. Deze valkuilen hebben betrekking op keuzes die de onderzoeker dient te maken. Het gaat om belangrijke keuzes met consequenties voor de inhoudelijke interpretatie van de resultaten. In deze bijdrage leggen we de klemtoon op de aard van deze problemen en trachten we deze eveneens te illustreren.

Een *eerste* en ongetwijfeld belangrijkste valkuil betreft de *teleenheid* die moet gekozen worden om de data te analyseren. De teleenheid is de statistische eenheid (unit of analysis) op basis waarvan uitspraken zullen worden gedaan. Parketdata bevatten echter uiteenlopende info. De onderzoeker dient hier een expliciete keuze te maken. Moeten feiten, processen verbaal (PV), betrokkenen (verdachten en slachtoffers) of tenlastenleggingen als teleenheid gekozen worden of zijn er nog andere mogelijkheden? Een *tweede* valkuil betreft de doorstroom naar de parketten van de gegevens die op het niveau van de politiediensten worden gevat. Zolang deze doorstroom niet op een geautomatiseerde wijze gebeurt, blijft er een onnodige foutenbron bestaan die invloed heeft op de accuraatheid van de parketdata. Een *derde* caveat betreft het vergelijken van parketten op basis van cijfergegevens. Zo