

Omgaan met sociale wenselijkheid: Inschatting van de dopingprevalentie aan de hand van de Randomized Response Technique

BERTRAND FINCOEUR^aWERNER PITTSCH^b

^a Senior onderzoeker, Institut des Sciences du Sport, Université de Lausanne (Corresp.: bertrand.fincoeur@unil.ch)

^b Docent, Sportwissenschaftliches Institut, Universität des Saarlandes

1. INLEIDING

Zowel in kwalitatief als in kwantitatief onderzoek naar gevoelige onderwerpen treedt vaak het probleem op van sociaal wenselijke antwoorden. Het eerlijk toegeven van wanpraktijken of controversiële meningen kan immers door de respondenten als bedreigend beschouwd worden. Terwijl iemand die aan een interview of een survey meewerkt de gevoeligste vragen kan overslaan, kan hij ook geneigd zijn om conventionele waarden te ondersteunen en om vooral algemeen geaccepteerde antwoorden te geven (KRUMPAL, 2013; LEE, 1993). In criminologisch onderzoek leidt sociale wenselijkheid dan vaak tot het minder frequent aangeven van het plegen van een delict (PAUWELS & PLEYSIER, 2005). Uiteraard is het wel mogelijk dat binnen bepaalde (subculturele) contexten de conventionele waarden “pro” illegale praktijken zijn maar, in het algemeen, verbergen de *insiders* voor de *outsiders* wat er in hun wereld gebeurt (BECKER, 1963). Als onderzoeker wordt deze situatie dus snel frustrerend. Ook voor beleidsmakers is het noodzakelijk om over betrouwbare informatie te beschikken. Inderdaad, hoe dient men onder deze omstandigheden de prevalentie in te schatten van een bepaald fenomeen waarvoor geen enkele andere methode toelaat om de werkelijke omvang vast te stellen? Zo vormt het probleem van sociale wenselijkheid eigenlijk een belangrijk knelpunt voor degene die niet tot de bestudeerde deviante subcultuur behoort maar die haar illegale praktijken tracht te evalueren aan de hand van *self-report* onderzoek (JUNGER-TAS & MARSHALL, 1999). De gevoeligheid van het thema heeft immers alle kans om een min of meer serieuze onderschatting van het fenomeen te veroorzaken. Dit probleem geldt uiteraard ook voor onderzoekers die proberen om dopinggebruik in kaart te brengen omdat het hier gaat om een verborgen fenomeen.

De pogingen om de dopingprevalentie in te schatten vormen een belangrijk aandachtspunt binnen de wetenschappelijke gemeenschap omdat er een brede consensus bestaat dat de cijfers en statistieken van de dopingcontroles geen betrouwbaar beeld van de dopingprevalentie geven (DIMEO & TAYLOR, 2013; LENTILLON-KAESTNER & OHL, 2011). Terwijl de *Wereld Antidoping Agentschap* het percentage van *Adverse Analytical Findings* rond 2% inschat¹, kan aangenomen worden dat de reële dopingprevalentie veel hoger is (PETROCZI, 2016; PITTSCH, 2011). In feite weerspiegelen de WADA-cijfers slechts het resultaat van de dopingcontroles (binnen-

1 Er kan een verschil worden gemaakt tussen *Adverse Analytical Findings* (AAF) en *Anti-Doping Rule Violations* (ADRV). De *Wereld Antidoping Code* definieert AAF als “a report from a laboratory or other WADA-approved entity that, consistent with the international standard for laboratories and related technical documents, identifies in a sample the presence of a prohibited substance or its metabolites or markers or evidence of the use of a prohibited method”. Niet iedere AAF leidt dus tot een sanctie omdat het gebruik van verboden producten en/of methoden kan toegelaten worden indien de atleet een *Therapeutic Use Exemption* heeft. Het percentage van dopinggevallen die gestraft kunnen worden is dus nog lager dan het percentage AAF.

en buiten-competitie, urine- en bloedcontroles) maar niet alle producten zijn bijvoorbeeld detecteerbaar. Daarom zou het percentage dopinggebruik in de topsport volgens onderzoek op basis van andere methoden (e.g. enquêtes bij atleten), eerder tussen 14 en 39% liggen (De HON, KUIPERS & VAN BOTTENBURG, 2015). PAOLI en DONATI (2014) bezorgden wel een methode om de grootte van de aanbodzijde van de Italiaanse dopingmarkt in te schatten maar dit geeft echter geen idee van de “reële”, maar tot nog toe onbekende, dopingprevalentie bij atleten op individueel niveau. Want niet alleen topsporters maar ook fitnessers, recreatieve sporters en zelfs niet-sporters gebruiken prestatieverbeterende producten en methoden om allerlei redenen, zoals levensstijl of gewichtsverlies (BOJSEN-MØLLER & CHRISTIANSEN, 2010; KRASKA, BUS-SARD & BRENT, 2010; VAN DE VEN, 2016).

In deze context hebben sommige onderzoekers geprobeerd om de risico's van sociaal wenselijke antwoorden te reduceren en, bijgevolg, om de betrouwbaarheid van de verzamelde data te verbeteren (PITSCH, 2016; STRIEGEL, ULRICH & SIMON, 2010). Het instrument dat in dit kader het vaakst gebruikt werd is de *Randomized Response Technique (RRT)* (LENSVELT-MULDERS, HOX, VAN DER HEIJDEN & MAAS, 2005; WARNER, 1965). Daarom bespreekt dit artikel deze onderzoeksmethode. Ten eerste wordt de RRT in algemene termen voorgesteld en uitgelegd. Ten tweede worden haar voordelen en beperkingen besproken dankzij enkele resultaten van een recent onderzoek dat deels aan de hand van deze methode werd uitgevoerd (FINCOEUR, 2016; FINCOEUR, FRENGER & PITSCH, 2014).

2. DE RANDOMIZED RESPONSE TECHNIQUE (RRT)

Het doel van de RRT is om het risico op sociaal wenselijk antwoordgedrag kleiner te maken dan bij een directe vraagstelling (WARNER, 1965). Dankzij deze methode kunnen de respondenten gevoelige vragen volledig anoniem beantwoorden. Bij directe bevraging zullen de respondenten inderdaad niet gauw toegeven zich schuldig te maken aan bijvoorbeeld dopinggebruik omdat ze mogelijk bang zijn voor de gevolgen van een “bekentenis” (PITSCH, 2011). De RRT beoogt om zo betrouwbaar mogelijk data te verzamelen aan de hand van een gerandomiseerd proces. Iedere respondent krijgt immers een kans om de bedreigende vraag ‘eerlijk’ te beantwoorden. Afhankelijk van het gekozen onderzoeksdesign, bepaalt de randomisatie welke vraag (gevoelig of niet-gevoelig) de respondent moet beantwoorden (*unrelated question randomized response model*: GREENBERG, ABUL-ELA & HORVITZ, 1969)² of dat de respondent standaard een bepaald antwoord moet geven, ongeacht de werkelijkheid (*forced answer technique*: BORUCH, 1971)³.

In beide gevallen, vertrekt de RRT van de hypothese dat respondenten meer geneigd zullen zijn om eerlijk te beantwoorden als ze overtuigd zijn dat hun antwoord op geen enkele manier geïdentificeerd zal kunnen worden. In ons onderzoek leidde de randomisatie bij een bepaald percentage van de respondenten tot een *forced answer*, ongeacht hun reële gedrag: het onderzoeksdesign is dus geacht om de anonimiteit van de respondenten te waarborgen. Inderdaad, dankzij het randomisatieproces, kan de onderzoeker helemaal niet achterhalen of de individuele respondent de bedreigende eigenschap al dan niet heeft.

2 Dit eerste type randomisatie werd bijvoorbeeld, toegepast in een criminologisch onderzoek door TRACY & FOX (1981).

3 In ons onderzoek gebruikten we dit tweede type randomisatie. Daarom focussen we in dit artikel alleen op dit type.

Ondanks de relatieve complexiteit van de methode worden de resultaten uit RRT-enquêtes vaak beschouwd als betrouwbaarder dan die van directe bevraging (COUTTS & JANN, 2011; ULRICH, SCHRÖTER, STRIEGEL & SIMON, 2012). Desalniettemin betekent een hogere prevalentie niet *per se* dat de resultaten betrouwbaarder zijn. De doelstelling is dus niet om zoveel mogelijk illegale feiten te vinden. De vergelijking tussen studies met en zonder RRT bewijst echter dat, hoewel er wellicht nog steeds onderschat wordt, RRT-onderzoek dichter bij de werkelijkheid komt (FIDLER & KLEINKLECHT, 1977; PITSCH & EMRICH, 2012; VAN DER HEIJDEN, VAN GILS, BOUTS & HOX, 2000).

Laat ons nu uitleggen hoe de techniek eigenlijk werkt. In onze enquête werd bijvoorbeeld de eerste gevoelige vraag zo gesteld: “Heeft u ooit bewust een verboden product of methode gebruikt teneinde uw sportprestaties te verbeteren?”. Deze vraag kan onbetwistbaar tot een sociaal geaccepteerd antwoord leiden. Om dit risico te verminderen, kregen de respondenten een bijkomende instructie. Een klassieke manier bestaat uit het opgooien van een muntstuk maar er bestaan veel verschillende randomisatietechnieken (COUTTS & JANN, 2011; SCHEERS, 1992; SHIMIZU & BONHAM, 1978). In ons onderzoek was het randomisatieproces gebaseerd op tien reeksen van elf cijfers (vb.: 29485266537). De respondenten werden dan gevraagd om één willekeurige reeks van cijfers te kiezen en te noteren. Ze konden ook het serienummer van een willekeurig eurobankbiljet kiezen. In beide gevallen werden echter alleen de 7^{de} tot 10^{de} cijfer gebruikt (vb. 29485266537) omdat deze cijfers gelijk verdeeld waren zowel in onze randomisatie als bij eurobankbiljetten.

Heeft u ooit bewust een verboden product of methode gebruikt teneinde uw sportprestaties te verbeteren (uitgezonderd TTN)?

Zijn er 7 dagen in een week?

- ☐ Ja
☐ Nee

De instructie is als volgt:

- Indien het derde cijfer startende vanaf het laatste cijfer van het serienummer van het bankbiljet of van de cijferreeks die u gekozen heeft een 0, 1 of 2 is, gelieve de vraag op de rechterkant te beantwoorden → De respondent zou dus “ja” moeten antwoorden aangezien iedere week 7 dagen heeft
- Indien het een ander cijfer is, gelieve de vraag op de linkerkant te beantwoorden → De respondent kan dus eerlijk antwoorden

Via het gebruik van wiskundige berekeningen, konden we op basis van het geheel van de ja-antwoorden (dus niet per persoon) berekenen hoeveel procent van alle ja-antwoorden tot de categorie van de linkervraag of rechtervraag behoort. We konden echter nooit achterhalen wie op welke vraag antwoordde.

De verdeling voor de twee eerste vragen was alleen het resultaat van een computergestuurde randomisatie:

- waarschijnlijkheid *forced answer* binnen RRT-groep 1 = 0.3
- waarschijnlijkheid *forced answer* binnen RRT-groep 2 = 0.7

De verdeling voor de derde en vierde vraag werd gemaakt op een manier dat de waarschijnlijkheid gebalanceerd was voor een respondent om een *honest answer* te kunnen geven. Het doel was immers om de waarschijnlijkheid te verminderen dat een respondent frustratie (door enkel *forced answers*) of achterdocht (door enkel *honest answers*) zou ervaren.

Door deze “*forced answer technique*” (BLAIR, IMAI & ZHOU, 2015) verzamelt de onderzoeker dus in principe twee soorten ‘ja’: wanneer de randomisatie de respondent ertoe leidt om ‘ja’ te

beantwoorden (*forced yes*) en wanneer de respondent eerlijk aangeeft ooit doping gebruikt te hebben (*honest yes*). In dit geval kan de onderzoeker onmogelijk weten of een bepaald 'ja-antwoord' het resultaat is van de randomisatie of de bekentenis van een wanpraktijk. Een nadeel van de RRT is dus dat het quasi uitsluitend gaat over het schatten van frequenties op collectief niveau. Door het randomisatieproces zijn er inderdaad geen individuele 'correcte' antwoorden meer, wat belangrijke implicaties heeft voor het bepalen van causale modellen, aangezien die hier alleen indirect bepaald kunnen worden, namelijk na vergelijking van de prevalentie tussen verschillende groepen. Deze beperking kan als een probleem gezien worden maar het is ook twijfelachtig dat causale modellen m.b.t. gevoelige thema's getoetst kunnen worden aan de hand van directe bevraging. De RRT levert op zijn minst een mogelijkheid om de implicaties te bespreken van causale individuele modellen die op collectief niveau aan het licht komen.

Ondanks de anonimiteit gebeurt het toch dat de respondenten de instructies niet volgen. Kortom er bestaan "valsspelers". Ten eerste kan een dopinggebruiker proberen zijn antwoord te verbergen omdat hij de techniek wantrouwt. In dit geval zal hij dus negatief antwoorden (*no cheater*) terwijl hij volgens het randomisatieproces (*forced yes*) of eerlijk (*honest yes*) had moeten antwoorden. Ten tweede kan een niet-dopinggebruiker, ondanks de randomisatie-instructies, bang zijn om zich schuldig te maken aan dopinggebruik (*forced yes*) hoewel hij in feite geen dopinggebruiker is (*honest no*). Ten derde zijn er achteloze respondenten, mensen die de techniek niet begrijpen en respondenten die opzettelijk de enquête afraffelen. Daarom hebben onderzoekers methoden ontwikkeld (CLARK & DESHARNAIS, 1998) en vervolgens getoetst (CRUYFF, VAN DEN HOUT, VAN DER HEIJDEN & BÖCKENHOLT, 2007; OSTAPCZUK, MUSCH & MOSHAGEN, 2011; SCHRÖTER, STUDZINSKI, DIETZ, ULRICH, STRIEGEL & SIMON, 2016) om het percentage van valsspelers in te schatten (*no cheater detection*). Deze methoden focussen op het valsspelend-gedrag waardoor de respondent een positief antwoord probeert te verbergen. Inderdaad is het zeldzamer om zichzelf zomaar te beschuldigen (i.e. positief beantwoorden zonder instructie van het randomisatieproces), hoewel dit ook mogelijk is bij saboteurs⁴. Het percentage dopinggebruikers ligt dus tussen het aantal respondenten die dopinggebruik toegeven (*honest dopers*) als ondergrens, en daarbij opgeteld het aantal valsspelers (*no cheaters*) als bovengrens.

In de praktijk dient de hele steekproef in twee delen opgesplitst te worden. Beide groepen moeten dan een verschillende kans ($p_1 \neq p_2$) krijgen om een *forced answer* te moeten geven (CLARK & DESHARNAIS, 1998; PITSCH, 2016). Via het randomisatieproces worden de respondenten van de eerste groep gestuurd naar de gevoelige vraag met een kans p_1 . De respondenten van de tweede groep hebben een kans p_2 ($1-p_1$). Concreet zijn er dus drie mogelijkheden. De *honest yes* respondenten (π) bevatten de atleten die dopinggebruik toegeven, indien het randomisatieproces hen vraagt om eerlijk te beantwoorden, én een deel van de *forced yes* antwoorden (de reële dopinggebruikers die positief beantwoorden want het randomisatieproces verplicht hen om zo te beantwoorden). De *honest no* respondenten (β) bevatten de niet-dopinggebruikers, die op de gevoelige vraag eerlijk negatief antwoorden, én een deel van de *forced yes* antwoorden (degenen die geen doping gebruiken maar aan wie het randomisatieproces vroeg om positief te antwoorden). De valsspelers (γ) bevatten de respondenten die de instructies niet volgen: ze beantwoorden altijd negatief, ongeacht het randomisatieproces of uit vrees eerlijk te antwoorden. Zo is $\pi + \beta + \gamma = 1$, i.e. de hele steekproef

4 FETH, FRENGER, PITSCH & SCHMELZEISEN (2014) hebben de methode uitgebreid met een *total cheater detection* (*yes & no cheaters*) maar de techniek vereist grotere steekproeven om betrouwbare resultaten te garanderen. In onze enquête werd deze methode dus niet toegepast.

(N). Om π , β en γ te kunnen inschatten, wordt N opgesplitst in twee delen (N_1 , N_2). Neem het onderstaande (theoretisch) geval:

TABEL 1: THEORETISCH VOORBEELD VAN RRT-ONDERZOEK

Groep	N_i	$p_i (p1 \neq p2)^5$	$y_i (\Sigma 'ja')$	$n_i (\Sigma 'nee')$
1	500	0.75	346	154
2	500	0.25	127	373

Binnen groep 1 hebben 500 respondenten drie kansen op vier (0.75) om een bepaald antwoord (*forced yes*) te moeten geven. In de praktijk antwoorden 346 respondenten positief terwijl 154 negatief antwoorden. Binnen groep 2 heeft de andere helft maar een kans op vier (0.25) om de gerandomiseerde vraag te krijgen, hoewel $\frac{3}{4}$ respondenten eerlijk op de gevoelige vraag mogen antwoorden. Het is dan ook logisch dat het negatieve antwoord het meest is gegeven in groep 2.

De verschillende percentages kunnen dan berekend worden aan de hand van de volgende formules (CLARK & DESHARNAIS, 1998: 163):

- $\pi = (p_2 y_1 / N_1 - p_1 y_2 / N_2) / (p_2 - p_1)$
- $\beta = (y_2 / N_2 - y_1 / N_1) / (p_2 - p_1)$
- $\gamma = 1 - (\pi + \beta)$

In het bovenstaande voorbeeld leidt dit tot de volgende resultaten:

- $\pi = 3,4\%$
- $\beta = 87,6\%$
- $\gamma = 9\%$

De prevalentie van het sociaal onwenselijke gedrag zou dan liggen tussen 3,4% en 12,4% ($\pi + \gamma$), in de onwaarschijnlijke hypothese waarin alle valsspelers 'leugenaars' zouden zijn. De RRT-techniek geeft dus geen zekerheid omdat niemand weet waarom een paar respondenten de instructies niet hebben gevolgd. Er kan dus alleen naar een marge worden verwezen.

Nu bezorgen we inhoudelijk een kort overzicht van enkele resultaten uit ons onderzoek. Daarna zullen we concluderen met de beperkingen van deze onderzoeksmethode.

3. DOPING IN DE WIELERSPORT IN KAART GEBRACHT

In het kader van een onderzoeksproject rond doping in de wielersport, hebben we een *online* survey uitgevoerd bij amateur competitieve wielrenners verbonden aan *Wielerbond Vlaanderen*. De survey had drie doelstellingen. Ten eerste, het inschatten van de dopingprevalentie. Ten tweede, het meten van de attitudes ten aanzien van doping en antidopingbeleid. Ten derde, het in kaart brengen van de evolutie van de (aanbodzijde van de) dopingmarkt (FINCOEUR, 2016; FINCOEUR, FRENGER & PITSCH, 2014). Hieronder rapporteren we alleen enkele resultaten met betrekking tot de dopingprevalentie omdat de andere twee delen op andere onderzoeksmethoden dan de RRT gebaseerd werden.

5 Door de codering van de enquête zijn deze proporties op voorhand gekend.

De enquête werd tussen mei en juli 2013 naar 2776 amateur competitieve wielrenners door-gemaid: 795 nieuwelingen (15-16 jaar oud), 711 aspiranten (17-18), 666 U23 (19-22) en 604 elites zonder contract (23+). De betrokkenheid van minderjarigen is gerechtvaardigd door studies die aandacht aan het dopinggebruik bij jongeren besteden (e.g. LAURE & BINSINGER, 2007; LORANG, CALLAHAN, CUMMINS, ACHAR & BROWN, 2011). In totaal hebben 767 respondenten de enquête ingevuld. De responsgraad was dus 27,6%. Gezien er gebruik werd gemaakt van een niet-probabilistische steekproeftrekking (alle leden van de vier categorieën ontvingen de online enquête), moet er dus gesteld worden dat een probleem van externe validiteit kan voorkomen (MAESSCHALCK, 2010).

De inschatting van de dopingprevalentie werd bestudeerd aan de hand van twee soorten gedrag en twee tijdschalen. We maakten een eerste onderscheid tussen het ooit gebruiken van een verboden product of methode (*lifetime prevalence*) en het gebruik in de afgelopen twaalf maanden (*last year prevalence*). Een tweede onderscheid werd gemaakt tussen het 'gewone' gebruik van een verboden product of methode en het gebruik zonder medische reden van een Toestemming wegens Therapeutische Noodzaak (TTN). Een atleet mag immers, volgens de antidopingregels en onder sommige voorwaarden, bepaalde verboden producten toch gebruiken maar een aantal atleten misbruiken deze TTN, als ze daarvoor geen medische reden hebben (BOURDON, SCHOCH, BROERS & KAYSER, 2014; OVERBYE & WAGNER, 2013). Onze enquête polste dus vier gedragingen (*lifetime doping*, *lifetime TTN*, *last year doping*, *last year TTN*). Hieronder worden de globale resultaten weergegeven.

TABEL 2: DOPINGPREVALENTIE (GLOBALE RESULTATEN, IN %) (N = 767)

	% Honest dopers	% No cheaters	% Honest no
<i>Lifetime doping</i>	5,3	34,4	60,3
<i>Last year doping</i>	0,4	26,3	73,3
<i>Lifetime TTN</i>	6,3	22,4	71,3
<i>Last year TTN</i>	1,7	23,3	75

Twee opmerkingen dienen geformuleerd te worden met betrekking tot de foutmarge van dit onderzoek. Ten eerste, beantwoordden niet alle respondenten onze vragen. Als de responsgraad lager is bij dopinggebruikers dan bij niet-dopinggebruikers – een plausibele hypothese – naderen onze resultaten misschien de ondergrens van de prevalentie maar dit is alleen maar speculatie. Ten tweede, moet de foutmarge geëvalueerd worden door *bootstrapping* (EFRON & TIBISHRANI, 1993). De resultaten met een 95%-betrouwbaarheidsinterval worden dan in de onderstaande tabel gepresenteerd.

TABEL 3: 95%-BETROUWBAARHEIDINTERVAL MET BC α -CORRECTION

	% Honest dopers	% No cheaters	% Honest no
<i>Lifetime doping</i>	0,0 – 13,5	29,1 – 40,7	50,8 – 68,1
<i>Last year doping</i>	0,0 – 8,9	20,8 – 30,7	68,4 – 79,6
<i>Lifetime TTN</i>	0,0 – 14,8	16,2 – 28,4	61,1 – 79,8
<i>Last year TTN</i>	0,0 – 9,7	17,4 – 29,4	67,7 – 82,7

Uit de resultaten van de survey blijken dus voornamelijk drie zaken.

Ten eerste, vrij voor de hand liggend, ligt het percentage respondenten dat toegeeft doping of een TTN zonder medische reden ooit gebruikt te hebben hoger dan het percentage tijdens de afgelopen twaalf maanden. Om de reële dopingprevalentie te kunnen inschatten, moet er echter een deel van de percentage van de *no cheaters* toegevoegd worden. Daarom ligt bijvoorbeeld het percentage respondenten dat een dopingproduct ooit gebruikt heeft tussen 5,3% (ondergrens) en 39,7% (bovengrens). Deze marge kan als eerder ruim beschouwd worden. Het dient echter vergeleken te worden met de meta-analyse VAN DE HON et al. (2015), die de dopingprevalentie, aan de hand van verschillende methoden, tussen 14 en 39% inschatte. Ook opmerkelijk is dat ons resultaat samenhangt met de percepties van de respondenten. Inderdaad, op de vraag “hoe hoog schat u het percentage gebruikers van verboden producten of methodes in uw sport?” gaven de respondenten ($N = 736$) een gemiddelde score van 31%. De percepties van de respondenten gaan dus niet veel hoger dan de bovengrens, hoewel dit waarschijnlijk ook bevestigt dat atleten geneigd zijn om de dopingprevalentie enigszins te overschatten (MORENTE-SANCHEZ & ZABALA, 2013).

Ten tweede formuleren we de hypothese dat het relatief lage percentage van *honest dopers*, hoewel hoger dan de officiële cijfers van de dopingcontroles, teruggekoppeld kan worden aan de evolutie van de wielercultuur ten aanzien van (anti)doping. Uit verschillende studies kan inderdaad afgeleid worden dat de jonge generaties van vandaag minder geneigd zijn om doping als een aanvaardbare gedragsmodaliteit te beschouwen (CHAN, HARDCASTLE, LENTILLON-KAESTNER, DONOVAN, DIMMOCK & HAGGER, 2014; OHL, FINCOEUR, LENTILLON-KAESTNER, DEFRANCE & BRISSONNEAU, 2015). Het lage percentage van *honest dopers* zou dus uitgelegd kunnen worden door het feit dat de tolerantiecultuur binnen de wielersport ten aanzien van doping vandaag minder sterk lijkt.

Ten derde, komt het misbruik van een TTN minstens even vaak voor dan het gewone gebruik van een verboden product of methode. Inderdaad heeft één respondent op zestien (6,3%; $N = 521$) zulk gedrag ooit vertoond. Zo bevestigt onze bevinding wat in de literatuur al werd aangehaald (BOURDON, SCHOCH, BROERS & KAYSER, 2014; OVERBYE & WAGNER, 2013). Wegens de nodige medeplichtigheid van gezondheidsdeskundigen in de dopingpraktijken (FINCOEUR, VAN DE VEN & MULROONEY, 2014; PAOLI & DONATI, 2014) dient dus het misbruik van TTN verder bestudeerd te worden.

4. CONCLUSIE

Het omgaan met sociale wenselijkheid is een klassiek probleem van kwantitatief onderzoek omdat respondenten niet snel zullen toegeven zich schuldig te maken aan gevoelige en/of illegale praktijken. Dit vormt onder andere een probleem voor onderzoekers die de prevalentie van een verborgen fenomeen proberen te schatten. In deze context werd de *randomized response technique* al lang gebruikt om dit probleem te omzeilen en om de betrouwbaarheid van de data te verhogen. Inderdaad heeft het gebruik van de RRT bewezen om tot betrouwbaardere resultaten te komen dan bij directe bevraging (LENSVELT-MULDERS, HOX, VAN DER HEIJDEN, & MAAS 2005; WOLTER, 2012). Het is dus opmerkelijk dat criminologen deze methode niet vaker gebruiken om illegale praktijken beter in kaart te brengen. Desalniettemin heeft de RRT een paar beperkingen.

In de eerste plaats kunnen de respondenten de RRT moeilijk vertrouwen. De ingewikkeldheid van de methode, het tijdverlies en de frustratie om soms niet te kunnen beantwoorden (indien de respondent een *forced answer* moet geven) kunnen inderdaad respondenten destabiliseren en/of demotiveren. Meer specifiek heeft het begrijpen van het randomisatie-

proces een impact op het vertrouwen in de techniek (De SCHRIJVER, 2012). Om gelijkaardige redenen heeft de RRT een negatieve impact op de responsgraad in vergelijking met klas-sieke bevraging (HOUSTON & TRAN, 2001). Een ander nadeel van de RRT is dat één vraag uit de vragenlijst maar over één gevoelig type van gedrag kan gaan. Om k eigenschappen te kunnen meten, net zoals met de *Unmatched Count Technique* (AHART & SACKETT, 2004) en de *Simple Sample Count* (PETROCZI et al., 2011), moet de techniek dan k keer herhaald worden. Er dienen dus k randomisatieprocessen ontwikkeld en gebruikt te worden. Dit betekent alweer een tijdverlies zowel voor de respondent als voor de onderzoeker. Daarom werd ook de *Randomized Count Technique* (RCT) recent ontwikkeld om tegelijk de prevalentie van verschillende gevoelige eigenschappen in te schatten (FINCOEUR, FRENGER & PITSCHE, 2014; FRENGER, PITSCHE & EMRICH, 2013).

Wegen dan de voordelen van *randomized response models*, en van RRT in het bijzonder, op tegen de nadelen? Er bestaat geen duidelijk antwoord. Zeker kunnen deze onderzoeks-methoden opportuniteiten opleveren om illegale praktijken in kaart te brengen, maar de bovenvermelde beperkingen maken het gebruik van RRT niet altijd volledig overtuigend. Bovendien moet de steekproef groot genoeg zijn in RRT-enquêtes. In het bijzonder betreffende het *cheater detection model*, is het van belang om grote steekproeven te hebben. Dit onderzoek was aan de ondergrens om de resultaten te kunnen exploiteren, en grotere steekproeven zijn zelfs noodzakelijk om een *total cheater detection* (yes- & no-cheaters) te kunnen toepassen. We bevelen toch aan om de RRT niet zonder *cheater detection* te gebruiken want alleen met deze methode kan een minimale en maximale prevalentie ingeschat worden. Een RRT-enquête zonder *cheater detection* bezorgt slechts de ondergrens van een prevalentie, i.e. een resultaat dat ook via directe bevraging bereikt kan worden (hoewel waarschijnlijk minder betrouwbaar). De onderzoeker moet dus een aantal afwegingen maken, namelijk over: het leveren van de nodige extra-inspanningen die vereist zijn om een RRT-survey uit te werken, zijn/haar kansen om een voldoende grote steekproef te hebben en de relevantie om over een (betrouwbaardere) onder- en bovengrens van de prevalentie van het bestudeerde fenomeen te beschikken. *Last but not least*, door deze eerder ingewikkelde methode, worden de resultaten van een RRT-survey vaak moeilijker communiceerbaar, voornamelijk in beleid georiënteerd onderzoek. Daarom, zonder de verdiensten van de RRT te negeren, zouden we misschien voor verder onderzoek de voorkeur geven aan recentere methoden, zoals de RCT. Op zijn minst stimuleren echter al deze technieken, RRT inbegrepen, theoretische en methodologische debatten binnen criminologisch onderzoek, waarvoor dit artikel een pleidooi wil maken.

REFERENTIES

- AHART, A.M., & SACKETT, P.R. (2004). A New Method of Examining Relationships between Individual Difference Measures and Sensitive Behavior Criteria. Evaluating the Unmatched Count Technique. *Organizational Research Methods*, 7, 101-114.
- BECKER, H.S. (1963). *Outsiders*. Glencoe: Free Press.
- BLAIR, G., IMAI, K., & ZHOU, Y. (2015). Design and Analysis of the Randomized Response Technique. *Journal of the American Statistical Association*, 110(511), 1304-1319.
- BOJSEN-MØLLER, J., & CHRISTIANSEN, A.V. (2010). Use of performance- and image-enhancing substances among recreational athletes: a quantitative analysis of inquiries submitted to the Danish anti-doping authorities. *Scandinavian Journal of Medicine & Science in Sports*, 20, 861-867.
- BORUCH, R. (1971). Assuring Confidentiality of Responses in Social Research: A Note on Strategies. *American Sociologist*, 6, 308-311.

- BOURDON, F., SCHOCH, L., BROERS, B., & KAYSER, B. (2014). French speaking athletes' experience and perception regarding the whereabouts reporting system and therapeutic use exemptions. *Performance Enhancement & Health*, 3(3/4), 153-158.
- CHAN, D., HARDCASTLE, S., LENTILLON-KAESTNER, V., DONOVAN, R.J., DIMMOCK, J.A., & HAGGER, M.S. (2014). Athletes' beliefs about and attitudes towards taking banned performance-enhancing substances: a qualitative study. *Sport, Exercise and Performance Psychology*, 3(4), 241-257.
- CLARK, S. J., & DESHARNAIS, R. A. (1998). Honest Answers to Embarrassing Questions: Detecting Cheating in the Randomized Response Model. *Psychological Methods*, 3, 160-168.
- COUTTS, E., & JANN, B. (2011). Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods & Research*, 40(1), 169-193.
- CRUYFF, M.J.L.F., VAN DEN HOUT, A., VAN DER HEIJDEN, P.G.M. & BÖCKENHOLT, U. (2007). Log-Linear Randomized-Response Models Taking Self-Protective Response Behavior Into Account. *Sociological Methods & Research*, 36(2), 266-282.
- DE HON, O., KUIPERS, H., & VAN BOTTENBURG, M. (2015). Prevalence of doping use in elite sports: a review of numbers and methods. *Sports Medicine*, 45(1), 57-69.
- DE SCHRIJVER, A. (2012). Sample survey on sensitive topics: Investigating respondents' understanding and trust in alternative versions of the randomized response technique. *Journal of Research Practice*, 8(1), 1-17.
- DIMEO, P. & TAYLOR, J. (2013). Monitoring drug use in sport: The contrast between official statistics and other evidence. *Drugs: education, prevention and policy*, 20(1), 40-47.
- EFRON, B., & TIBSHIRANI, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- FETH, S., FRENGER, M., PITSCH, W., & SCHMELZEISEN, P. (2014). *Cheater-Detection bei der Randomized Response-Technik. Herleitung, Analyse und Anwendung* (Schriften des Europäischen Instituts für Sozioökonomie e.V., 8). Saarbrücken: universaar. Online available https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=oahUKEWjnzYG7n8zUAhWG8RQKHbPDBAKQFggsMAA&url=http%3A%2F%2Funiversaar.uni-saarland.de%2Fmonographien%2Fvolltexte%2F2015%2F134%2Fpdf%2FRRTMitCheaterDetection_ebook.pdf&usg=AFQjCNHtOHelv8BuE1aQxe2xFImDM9iUg
- FIDLER, D.S. & KLEINKNECHT, R.E. (1977). Randomized response versus direct questioning. Two data-collection methods for sensitive information. *Psychological Bulletin*, 84(5), 1045-1049.
- FINCOEUR, B. (2016). *Le marché du dopage dans le cyclisme sur route belge et français: analyse de la demande, de l'offre et de l'impact de la lutte antidopage*. Doctoraatsthesis in criminologie: KU Leuven.
- FINCOEUR, B., FRENGER, M., & PITSCH, W. (2014). Does one play with the athletes' health in the name of ethics? *Performance Enhancement & Health*, 2(4), 182-193.
- FINCOEUR, B., VAN DE VEN, K., & MULROONEY, K. (2014). The Symbiotic Evolution of Anti-Doping and Supply Chains of Doping Substances: How Criminal Networks May Benefit From Anti-Doping Policy. *Trends in Organized Crime*, 17, 1-22.
- FRENGER, M., PITSCH, W., & EMRICH, E. (2013). Ehrliche Antworten auf sensitive Fragen?! – Entwicklung einer Befragungsmethode und erste Anwendungsplanung. In H. KEMPF, S. NAGEL, H. DIETL (Eds.), *Im Schatten der Sportwirtschaft* (pp. 241-252). Schorndorf: Hofmann.
- GREENBERG, B., ABUL-ELA, A., & HORVITZ, D. (1969). The Unrelated Question Randomized Response Model: Theoretical Framework. *Journal of the American Statistical Association*, 64(326), 520-539.
- HOUSTON, J., & TRAN, A. (2001). A survey of tax evasion using the randomized response technique. *Advances in Taxation*, 13, 69-94.
- JUNGER-TAS, J., & MARSHALL, I.H. (1999). The Self-Report Methodology in Crime Research. *Crime and Justice*, 25, 291-367.

- KRASKA, P., BUSSARD, C., & BRENT, J. (2010). Trafficking in Bodily Perfection: Examining the Late-Modern Steroid Marketplace and its Criminalization. *Justice Quarterly*, 27(2), 159-185.
- KRUMPAL, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47(4), 2025-2047.
- LAURE, P., & BINSINGER, C. (2007). Doping prevalence among preadolescent athletes: a 4-year follow-up. *British Journal of Sports Medicine*, 41(10), 660-663.
- LEE, R.M. (1993). *Doing Research on Sensitive Topics*. London: Sage.
- LENSVELT-MULDERS, G., HOX, J., VAN DER HEIJDEN, P., & MAAS, C. (2005). Meta-Analysis of Randomized Response Research. Thirty-Five Years of Validation. *Sociological Methods & Research*, 33, 315-348.
- LENTILLON-KAESTNER, V., & OHL, F. (2011). Can we measure accurately the prevalence of doping? *Scandinavian Journal of Medicine & Science in Sports*, 21(6), 132-142.
- LORANG, M., CALLAHAN, B., CUMMINS, K., ACHAR, S., & BROWN, S. (2011). Anabolic Androgenic Steroid Use in Teens: Prevalence, Demographics, and Perception of Effects. *Journal of Child & Adolescent Substance Abuse*, 20(4), 358-369.
- MAESSCHALCK, J. (2010). Methodologische kwaliteit in het kwalitatief criminologisch onderzoek. In T. DECORTE, D. ZAITCH (Eds.), *Kwalitatieve methoden en technieken in de criminologie* (pp. 120-145). Leuven/Den Haag: Acco.
- MORENTE-SANCHEZ, J., & ZABALA, M. (2013). Doping in Sport: A Review of Elite Athletes' Attitudes, Beliefs, and Knowledge. *Sports Medicine*, 43(6), 395-411.
- OHL, F., FINCOEUR, B., LENTILLON-KAESTNER, V., DEFANCE, J., & BRISSONNEAU, C. (2015). The Socialization of Young Cyclists and the Culture of Doping. *International Review for the Sociology of Sport*, 50(7), 865-882.
- OSTAPCZUK, M., MUSCH, J., & MOSHAGEN, M. (2011). Improving self-report measures of medication non-adherence using a cheating detection extension of the randomised-response-technique. *Statistical Methods in Medical Research*, 20(5), 489-503.
- OVERBYE, M., & WAGNER, U. (2013). Between medical treatment and performance enhancement: An investigation of how elite athletes experience Therapeutic Use Exemptions. *International Journal of Drug Policy*, 24(6), 579-588.
- PAOLI, L., & DONATI, A. (2014). *The Sports Doping Market. Understanding Supply and Demand, and the Challenges of Their Control*. New York: Springer.
- PAUWELS, L., & PLEYSIER, S. (2005). Effecten van antwoordstijlen in etiologisch self-report-onderzoek. Een 'causale modellenbenadering'. *Tijdschrift voor Criminologie*, 47(1), 42-61.
- PETROCZI, A. (2016). Indirect measures in doping behavior research. In V. BARKOUKIS, L. LAZURAS, H. TSORBATZOUKIS (Eds.), *The Psychology of Doping in Sport* (pp. 93-110). London: Routledge.
- PETROCZI, A., NEPUZ, T., CROSS, P., TAFT, H., SHAH, S., DESHMUKH, N., SCHAFFER, J., SHANE, M., ADESANWO, C., BARKER, J., & NAUGHTON, D. P. (2011). New non-randomised model to assess the prevalence of discriminating behaviour: a pilot study on mephedrone. *Substance Abuse, Treatment, Prevention and Policy*, 6(1), 20.
- PITSCH, W. (2016). Minimizing response bias: an application of the randomized response technique. In V. BARKOUKIS, L. LAZURAS, H. TSORBATZOUKIS (Eds.), *The Psychology of Doping in Sport* (pp. 111-125). London: Routledge.
- PITSCH, W. (2011). Doping controls between test theory and ethics – unintended consequences of in principle imperfect doping tests. In E. EMRICH, W. PITSCH (Eds.), *Sport and Doping. The Analysis of an Antagonistic Symbiosis* (pp. 101-113). Frankfurt: Peter Lang.
- PITSCH, W., & EMRICH, E. (2012). The frequency of doping in elite sport: Results of a replication study. *International Review for the Sociology of Sport*, 47(5), 559-580.
- SCHEERS, J. (1992). Methods, Plainly Speaking. A Review of Randomized Response Techniques. *Measurement and Evaluation in Counselling and Development*, 25(1), 27-41.

- SCHRÖTER, H., STUDZINSKI, B., DIETZ, P., ULRICH, R., STRIEGEL, H., & SIMON, P. (2016). A Comparison of the Cheater Detection and the Unrelated Question Models: A Randomized Response Survey on Physical and Cognitive Doping in Recreational Triathletes. *PLOS One*, 11(5), 1-11.
- SHIMIZU, I.M., & BONHAM, G. (1978). Randomized Response Technique in a National Survey. *Journal of the American Statistical Association*, 73, 35-39.
- STRIEGEL, H., ULRICH, R., & SIMON, P. (2010). Randomized Response Estimates for Doping and Illicit Drug Use in Elite Athletes. *Drug and Alcohol Dependence*, 106, 230-232.
- TRACY, P.E., & FOX, J.A. (1981). The Validity of Randomized Response for Sensitive Measurements. *American Sociological Review*, 46(2), 187-200.
- ULRICH, R., SCHRÖTER, H., STRIEGEL, H., & SIMON, P. (2012). Asking sensitive questions: A statistical power analysis of randomized response models. *Psychological Methods*, 17(4), 623-641.
- VAN DER HEIJDEN, P., VAN GILS, G., BOUTS, J., & HOX, J. (2000). A comparison of randomized response, computer-assisted self- interview, and face-to-face direct questioning – Eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods & Research*, 28, 505-537.
- VAN DE VEN, K. (2016). *The Formation and Development of Illicit Performance and Image Enhancing Drug Markets: Exploring Supply and Demand, and Control Policies in Belgium and the Netherlands*. University of Kent: Policy report.
- WARNER, S.L. (1965). Randomized-response. A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, 63-69.
- WOLTER, F. (2012). *Heikle Fragen in Interviews*. Wiesbaden: VS.