Wetenschappelijk Onderzoek- en Documentatiecentrum
Ministerie van Veiligheid en Justitie

Onderzoeksafdeling Rechtsbestel, Wetgeving, Internationale- en vreemdelingen-aangelegenheden (RWI)

Turfmarkt 147
2511 DP Den Haag
Postbus 20301
2500 EH Den Haag
www.wodc.nl

Contactpersoon
r

Notitie

Reactie WODC op Review APE van de Korte-termijn Monitor Raadsman bij verhoor (Significant, 2016)

Datum
7 november 2016

Projectnaam
Monitor raadsman bij verhoor

Ons kenmerk
2000381

In het kader van de opgestarte 'Lange-termijn Monitor' Raadsman bij verhoor heeft het WODC een reactie opgesteld op recente kanttekeningen door APE – in een review i.o.v. de NoVA – bij de eerdere 'Korte-termijn monitor' van Significant (juli 2016)[1] en het verweer van Significant op die kritiek. Op 11 oktober 2016 is deze reactie aan de Directie Rechtsbestel verstrekt. Hieronder geven we de letterlijke WODC-reactie weer, gevolgd door een weergave van de respons hierop van Significant op 19 oktober 2016. Besloten wordt met de conclusie tot welke het WODC uiteindelijk op 20 oktober is gekomen.

Reactie WODC (11 oktober 2016)
Onze puntsgewijze reactie volgt na een korte introductie op de Significant-monitor en de onderliggende gegevens. De reactie wordt besloten met een samenvattende conclusie en een korte vooruitblik op de 'Lange-termijn monitor', uit te voeren door het WODC.

- In de 'Korte-termijn monitor' worden conclusies getrokken over:
  - het gemiddelde aantal verhoren per zaak op jaarbasis (alleen piketfase);
  - waarbij tijdens het verhoor al of niet rechtsbijstand is verleend; en
  - de gemiddelde verhoorduur – ook uitgesplitst naar wel/geen verhoorbijstand – over de 1,5 maandsperiode na de invoering per 1 maart 2016.

- Wat de politie betreft, worden hiervoor bestanden voortkomend uit drie registratiesystemen gebruikt: BVH ofwel het basisregistratiesysteem; SUMM-IT, voor zwaardere/gevoelige (recherche-)zaken; en AVR, een klein deelsysteem op basis van audio/visueel geregistreerde verhoren.
- De vraag is, of Significant op basis van deze drie registratiebronnen en (al) hun beperkingen verantwoorde conclusies heeft kunnen trekken.
- Significant geeft steeds een beknopt overzicht van beperkingen en knelpunten in de politieregistraties en hoe de onderzoekers daarmee in het bewerkings- en analyseproces zijn omgegaan. Belangrijkste knelpunten zijn volgens de onderzoekers zelf:

[1] De eerste monitor van het per 1 maart 2016 geldende recht op verhoorbijstand voor meerderjarige verdachten.

- In sommige systemen ontbreekt informatie over kern-variabelen zoals eindtijden verhoor (bijv. in BVH maar voor 23% geregistreerd) en over de aanwezigheid van de raadsman (j/n). De laatste variabele is daarom niet ontleend aan de systeemregistraties in BVH, maar handmatig op basis van 3.779 PV's-verhoor gereconstrueerd.
- De overlap tussen BVH en SUMM-IT is onbekend. Dit omdat niet wordt gewerkt met een gemeenschappelijk zaaknummer of -ID in de verschillende registratiesystemen. Wel is duidelijk dat SUMM-IT complexere, 'recherche'-achtige zaken omvat.[2] De kleine set van AVR-registraties betreft veelal ook zwaardere, specifieke typen zaken[3] die volgens Significant waarschijnlijk ook door BVH zijn afgedekt. Omdat AVR echter een grofmazig inzicht biedt in de verhoorduur van dit type zaken, is deze informatie toch gebruikt om witte vlekken in BVH enigszins te compenseren.

- In de APE-review worden een aantal knelpunten in de Korte-termijn monitor naar voren gebracht. Een gedeelte daarvan betreft gebreken in de onderliggende (politie)data zelf, een ander gedeelte de manier waarop Significant die data heeft bewerkt en geanalyseerd om tot conclusies te komen over de gemiddelde verhoorduur, met/zonder raadsman.

Review APE	**'Verweer' Significant**	**Onze reactie**
BVH volgens APE van onvoldoende kwaliteit en daardoor niet betrouwbaar; met name vanwege slechts voor 23% ingevulde verhoor-eindtijden. Om deze reden had Significant de BVH-verhoren (over 1,5 maand) niet mogen extrapoleren naar jaarbasis.	De uitval van eindtijden is volgens het verweer van Significant niet selectief, er is dus geen systematische 'bias' in termen van zaakzwaarte. Voor een goed totaalbeeld is extrapolatie naar jaarbasis volgens Significant essentieel en kan niet worden volstaan met een gemiddelde op basis van alleen de beschikbare gegevens.	Wat Significant in het verweer opmerkt had in de oorspronkelijke rapportage van juli 2016 direct moeten worden verduidelijkt. Afwezigheid van systematische bias is niettemin een overtuigend argument om in de extrapolatie naar jaarbasis met alle BVH-verhoren te werken. *Wel rest nog de vraag of de bias echt verwaarloosbaar is. Immers, bij

[2] De SUMM-IT gegevens heeft men maar voor enkele regio's boven water kunnen krijgen (vnl. Oost-Nederland) en uitspraken die men hieraan ontleent, zijn dan ook niet generaliseerbaar – hoewel volgens Significant vanuit de politie geen redenen zijn genoemd om te verwachten dat de cijfers in andere regio's sterk zullen afwijken.

[3] vgl. voor criteria voor AVR: OM Aanwijzing AVR, 2010A018gp.

		extrapolatie wordt deze uitvergroot. Ook is nog de vraag of er geen 'verborgen bias' is (bijv. niet-registreren als gevolg van bepaalde verhoorsituaties).
Ongeloofwaardig dat slechts voor 27% een raadsman aanwezig is geweest – er is sinds 1 maart immers een beleidsregel	Dat de praktijk niet overeenkomt met de OM-beleidsregel is geen argument om registraties te wantrouwen. Dat de raadsman aanwezig is geweest of niet, is voorts een sleutelvariabele op basis waarvan verhoorduren kunnen worden vergeleken en de verschillen getoetst op significantie (wat overigens niet het geval bleek te zijn)..	Het verweer snijdt hout. Het gaat bovendien om een recht (ipv plicht) voor de verdachte. In de Lange-Termijn Monitor zal aan de hand van PV's en interviews opnieuw worden bestudeerd wat het percentage is en hoe dit begrepen kan worden. APE lijkt overigens niet te onderkennen dat dit percentage handmatig is ontleend aan 3.779 PV's en niet aan BVH.
Review APE	**'Verweer' Significant**	**Onze reactie**
Verhoren zijn in het proces van ontdubbeling van registratiegegevens onterecht verwijderd	Dit is volgens Significant (in rapport en verweer) niet het geval; ontdubbeling is zo uitgevoerd dat de totale verhoorduur van overlappende verhoren als één verhoor is gerekend. (Op p.8 van het originele Significant-rapport staat voorts dat 406 vervolg-verhoren starten voordat het voorafgaande verhoor is afgelopen. 'Van 377 van die overlappende vervolgverhoren is geen verhoorduur bekend en die zijn verwijderd' - lees: met instandhouding van de	Significant lijkt het BVH-bestand op een verantwoorde manier te hebben geschoond van duidelijke dan wel plausibele registratie-fouten. Dit had in de Korte-termijn monitor echter meer toelichting/ nadruk kunnen krijgen. Significant experimenteert op

	duur van het eerst-geregistreerde verhoor). In de Korte-termijn monitor en tevens in het verweer staat ten slotte dat IVS-verhoren niet zijn verwijderd en dus meetellen in het aantal verhoren per zaak, en daarmee in de totale verhoorduur.	p. 4 met het weglaten van zeer korte en zeer lange verhoren - het gemiddelde blijkt niet instabiel. De verwijderde overlappende *vervolg*verhoren omvatten overigens niet meer dan 1,4% van de 26.408 registraties.
Review APE	**'Verweer' Significant**	**Onze reactie**
Verhoren overlappen, waardoor wat in feite 1 lang verhoor is, staat geregistreerd als meerdere korte verhoren	Duidelijk overlappende verhoren zijn als 1 verhoor meegenomen, met de eindtijd van het tweede.	Waar mogelijk is de langste verhoorduur meegenomen. Opnieuw gaat het om zeer beperkte aantallen in relatie tot het totaal van 26.408 in BVH geregistreerde verhoren.
Het is niet duidelijk of de verhoorduur inclusief dan wel exclusief nalopen van het PV-verhoor (evt. met raadsman) is.	Dit is een registratieprobleem dat Significant ook signaleert (p. 19 rapport) en voorstelt te verhelderen in de Lange-termijn Monitor.	In de Lange-termijn Monitor zal worden gepoogd hier meer vat op te krijgen.
Gelet op de beperkingen aan BVH (behalve uitval eindtijden ook: verschillen gemiddelde verhoorduur A/B met SUMM-IT en AVR) dient volgens APE niet een totaalgemiddelde genomen te worden als bovengrens, maar AVR. APE berekende, op basis van de geaggregeerde gegevens in het Significant-rapport nieuwe bandbreedtes rondom de gemiddelde verhoorduur. Als gevolg van de (andere) aannames die APE maakt, met AVR als leidende bron, wordt de bovengrens aanzienlijk opgerekt.	Significant stelt dat BVH de beste bron is; dit systeem dekt de grootste stroom aan verhoren af. SUMM-IT en AVR zijn zeer specifieke systemen die voor een subset aan zaken worden gebruikt. De overlap tussen de systemen is onbekend. 'Daarom is gewerkt met scenario's van volledige overlap (c.q. enkel BVH is informatief, SUMM-IT-informatie zit reeds in BVH) tot SUMM-IT als volledige aanvulling (c.q. SUMM-IT en BVH geven samen het totaalbeeld)'. Zie ook p. 18 van de Korte-	Het verweer snijdt zeker hout. APE lijkt er ten onrechte van uit te gaan dat AVR in enige mate representatief is voor de hele stroom aan zaken. Waarom zou men het gemiddelde van alléén een kleine subset van zaken, die niet-representatief zijn voor de werkelijkheid, als bovengrens voor alle zaken willen nemen?

	termijn Monitor.	Niettemin bevatten SUMM-IT en AVR relatief meer informatie over bijv. de verhoorduur en de aanwezigheid van raadsman. Het is dan ook goed dat analyses van deze systemen eerst ook afzonderlijk in de Korte-termijnmonitor aan bod komen.
Er is sprake van inconsistente extrapolatie van aantallen verhoren naar jaarbasis tussen de verschillende bronnen	De extrapolatie-keuzes worden deels verantwoord in de rapportage; maar pas in het verweer stelt Significant dat, waar bekend, de aantallen uit 2015 zijn gebruikt – en waar niet, is geëxtrapoleerd van 1,5 maand naar jaarbasis.	In de Korte-termijn monitor staat nog niet vermeld wat in het verweer wel staat. Het originele rapport ontbeert dus een stukje verantwoording achter de door APE gevonden 'inconsistenties'.

Concluderend:

- De analyses door Significant van aantal en duur van de verhoren zijn gebaseerd op drie verschillende systemen, 'waarbij elk systeem voor een (groter of kleiner) deel van het geheel informatie bevat', aldus Significant in het verweer. Significant stelt hierin voorts dat BVH 'de beste bron' is. Wellicht is het beter te spreken van de bron met de minste gebreken. Hoewel niet representatief voor zwaardere zaken dekt BVH de grootste stroom aan verhoren af.
- Significant hanteert expliciete en veilige aannames c.q. scenario's voor de bandbreedtes rond de gemiddelde verhoorduur, namelijk dat de registratiesystemen volledig (=ondergrens o.b.v. BVH) dan wel helemaal niet (=bovengrens, o.b.v. het totaal verhoren in BVH, SUMMIT en AVR) overlappen. Dat APE alleen AVR wil inzetten om tot een nieuwe bovengrens te komen, is niet realistisch omdat dit systeem een zeer specifieke subset van zaken omvat, waarvan het bovendien aannemelijk is dat deze ook in BVH voorkomen.
- Voorts is de driekwart registratie-uitval van verhoorduren – althans, volgens het verweer van Significant – zonder 'bias', ofwel: er is niet relatief veel uitval van lichtere dan wel zwaardere zaken. Dit staat echter niet expliciet in de originele Korte-termijn monitor (bijv. in Tabel 7 valt dit niet te zien). Als het feitelijk klopt, is de afwezigheid van selectieve uitval een overtuigend argument. Het zou extrapolatie van bevindingen naar jaarbasis inderdaad toestaan – mits de vertekening echt nihil is, anders zou die worden uitvergroot.

- Al met al heeft Significant geroeid met de op dat moment meest geschikte riemen en gevolgtrekkingen goeddeels verantwoord. Goeddeels, want op twee niet onbelangrijke punten, waarop APE terecht aanslaat, ontbreekt toelichting: (1) De volgens het verweer niet-selectieve uitval van de eindtijden verhoor in BVH. (2) De toelichting bij de extrapolatie van 1,5 maand naar jaarbasis op p.17; deze is pas met het verweer erbij compleet. De verslaglegging en verantwoording van de keuzes in de Korte-termijn monitor had ook als geheel uitgebreider gekund – met name wat betreft de gemaakte aannames om uiteindelijk tot de bandbreedtes rond de verhoorduur te komen. Zo omvat de conclusieparagraaf slechts 2 regels. De beknoptheid van de verantwoording is het vertrouwen in de validiteit van de Significant-rapportage waarschijnlijk niet ten goede gekomen.

Ten slotte:

- De *Lange*-termijn Monitor, die verder reikt dan de piketfase, is een onderzoek met een eigen, opnieuw te verantwoorden aanpak. Deze dient ter beantwoording van vier onderzoeksvragen die verder reiken dan alleen de verhoorduur. Deze onderzoeksvragen behelzen ook een procesevaluatie van de organisatie en de invoering van het recht, en wat het voor de betrokken organisaties in de praktijk betekent om de aanwezigheid van advocaten bij verhoren mogelijk te maken. Ten slotte wordt de feitelijke toepassing en invulling van het recht tijdens de verhoren in de Lange-termijn monitor betrokken. Belangrijke ambities zijn om:
    - Een ruimere onderzoeksperiode af te dekken voor zowel registraties in BVH als SUMM-IT. Voorts zal energie worden gestoken in een hogere (geografische) representativiteit van SUMM-IT-registraties.
    - Meervoudige dataverzamelingsmethoden: naast registraties ook dossieronderzoek (steekproeven PV's-verhoor), en voorts in elk geval interviews. De dossiers en de interviews fungeren ter aanvulling en tevens ter check op registratielacunes.
    - Zo volledig mogelijk rapporteren over de gevolgde aanpak (dataverzamelings- en analysemethoden), inclusief de verantwoording van gemaakte keuzes. Dit gebeurt onder supervisie van een wetenschappelijk georiënteerde begeleidingscommissie.

Respons Significant (19 oktober 2016)

Significant gaf in respons hierop aan de analyse van het WODC op prijs te stellen, en de suggestie om nog wat explicieter te verantwoorden goed te kunnen volgen. De onderzoekers lichtten daarop volgend een aantal zaken nader toe, met name hoe om is gegaan met mogelijke bias van verhoorregistratie in het BVH bestand. "Als aanvullende toelichting naar aanleiding van de reactie van het WODC het volgende: wij beweren niet dat er van geen enkele bias sprake is. Zoals het WODC terecht opmerkt is een verborgen bias niet vast te stellen. Wat wij wel steeds hebben gedaan is een bias in de zin van mogelijk structurele *onderschatting* (wat APE suggereert) van de berekende gemiddelde verhoorduur vermijden. Het vermijden van *onderschatting* heeft op onderdelen zelfs geleid tot een (bewuste) *overschatting* van de verhoorduur.
Het voorkómen van onderschatting hebben wij op de volgende manieren gedaan:

- Er is wel degelijk verschil in de registratie van verhoorduur naar zwaarte van de zaak of verhoor in BVH (zie de tabel hieronder, de aantallen staan

Onderzoeksafdeling Rechtsbestel, Wetgeving, Internationale- en vreemdelingen-aangelegenheden (RWI)

Datum
5 oktober 2016

Ons kenmerk
2000381

overigens ook in de rapportage, maar verdeeld in tabel 1 en 3). We hebben daar bewust naar gekeken en rekening mee gehouden.

- Mede daarom is in de berekeningen van de verhoorduur steeds een uitsplitsing gemaakt naar type zaak. Met name voor het onderscheid naar categorie A- en B-zaken is dat relevant, omdat daarbij ook de verhoorduur substantieel verschilt. Binnen de categorie B-zaken valt het verschil weg in de afronding (zowel met als zonder weging voor het aantal zaken is de gemiddelde duur 49 minuten). We rapporteren consequent op het niveau van categorie A- en B-zaken. Dit verschil in registratie in BVH levert daarom geen vertekening op voor de conclusies over de gemiddelde verhoorduur binnen de categorieën.

BVH: beschikbaarheid van verhoorduur			
Categorie feit	Zaken BVH	Duur eerste verhoor bekend	% verhoor duur bekend
A: 12 jaar plus	908	189	21%
B: 4 tot 12 jaar	21.117	5.591	26%
B: 6 tot 12 jaar	3.126	701	22%
B: Overig VH	17.991	4.890	27%
C: Niet VH	3.898	1.376	35%
Onbekend	485	168	35%

- Van de IVS verhoren wordt nooit de verhoorduur geregistreerd in BVH (dit betreft een belangrijk deel van de verhoren). Deze verhoren duren volgens onze informatie in de praktijk (zeer) kort. Wij hebben er desondanks voor gekozen deze verhoren wel mee te rekenen in de aantallen verhoren, als waren het doorsnee verhoren met een gemiddelde duur. Voor deze IVS verhoren is daardoor vrijwel zeker sprake van een overschatting van de werkelijke verhoorduur.

- Voor zwaardere zaken is aanvullend gebruik gemaakt van de verhoorduren voor zover bekend op basis van AVR en SummIT om een bovengrens scenario vast te stellen. Er is echter zeker sprake van grote overlap tussen BVH, AVR en SummIT. Door in de berekening geen overlap te veronderstellen is in het 'bovengrens scenario' sprake van een overschatting. De berekende gemiddelden in het 'bovengrens scenario' zijn in de verhoudingen van zaakzwaarte en aantallen op jaarbasis gewogen om tot een gecorrigeerd totaal gemiddelde per categorie (A- of B-zaken) te komen. Dat is nodig omdat anders door verschillende beschikbaarheid van informatie en periodes uit verschillende bronnen de verhoudingen niet corresponderen met werkelijke verhoudingen in jaarlijkse aantallen. Zonder weging zou BVH te weinig meewegen en AVR te zwaar (zie ook onderstaande tabel voor een overzicht van verschil in gewichten op basis van onderliggende tabellen in de rapportage).

Onderzoeksafdeling Rechtsbestel, Wetgeving, Internationale- en vreemdelingen-aangelegenheden (RWI)

Datum
5 oktober 2016

Ons kenmerk
2000381

Categorie feit	Bron	Schatting verhoren piketfase (jaarbasis)	%	Aantal verhoorduren bekend	%
A: 12 jaar plus	BVH	13.593	73%	908	47%
	AVR	3.741	20%	889	46%
	Summit	1.304	7%	144	7%
	Subtotaal	18.638	100%	1.941	100%
B: overig VH	BVH (6 tot 12 jaar)	48.171	17%	701	11%
	BVH (overig VH)	231.790	82%	4.890	74%
	AVR	2.955	1%	1.010	15%
	Summit	181	0%	20	0%
	Subtotaal	283.098	100%	6.621	100%

Daarbij nog de volgende opmerkingen:

- Voor BVH gegevens is geen extrapolatie toegepast, maar waren werkelijke aantallen zaken per categorie over 2015 bij ons bekend (het gebruikte jaartal had inderdaad in de rapportage genoemd moeten worden). In tabel 21 in de rapportage is aangegeven voor welke onderdelen daarvoor gebruik is gemaakt van een extrapolatie om tot jaarcijfers te komen. Het totaal aantal verhoren op jaarbasis in SummIT is benaderd op basis van: 230 verhoren in 51 dagen in Oost Nederland. Oost-Nederland omvat ongeveer 16% van het totaal aantal zaken op basis van verhoudingen in BVH. Verdeling over A en B feiten en deel in piketfase is conform bevindingen in de beschikbare data. Op basis van uit de data afgeleide informatie over het percentage verhoren in piketfase (per bron en categorie feit) is een schatting gemaakt van het jaarlijks aantal verhoren in piketfase. Deze schatting is als weegfactor gehanteerd.
- De berekende gewogen gemiddelden zijn geen exacte werkelijkheid, maar een benadering met onzekerheden. Op basis van zowel aantal verhoren per zaak als verhoorduren zijn statistische (95%) betrouwbaarheidsintervallen berekend voor de gemiddelden (zie ook tabel 24). Om een interval te bepalen voor de totale verhoorduur in de piketfase zijn de uitersten van deze intervallen met elkaar vermenigvuldigd. Hiermee is een ruim interval gehanteerd, dat breder is dan een 95% betrouwbaarheidsinterval. Ook hier is dus bij het berekenen van de bovengrens ruim gerekend en eerder sprake van overschatting dan van onderschatting.

Samenvattend: wij zijn er in de uitvoering van het onderzoek steeds alert op geweest niet uit te komen op een onderschatting van de verhoorduur door een mogelijke bias of ontbrekende gegevens. Mede om die reden zijn de berekende cijfers steeds per categorie van feiten apart gepresenteerd. Op onderdelen is hierin zelfs bewust een overschatting toegelaten (door IVS verhoren mee te rekenen als waren het reguliere verhoren; het gebruik van overlappende registraties in AVR en SummIT met een passende weging naar aantallen op jaarbasis voor deze bronnen). Ten slotte is door statistische intervallen inzichtelijk gemaakt wat lagere en hogere uitkomsten van de berekeningen zouden kunnen zijn wanneer meer gegevens van vergelijkbare aard worden verkregen."

Conclusie WODC

Deze uitgebreide reactie van Significant biedt, samen met het eerdere 'verweer' op de APE-review en de oorspronkelijke Korte-termijnmonitor genoeg basis om te zeggen: De op grond van de beschikbare gegevens getrokken conclusies hebben voldoende marge en voorzichtigheid en zijn voldoende onderbouwd. Met behulp

van ruime aannames is verantwoord omgegaan met de tekortkomingen in het onderliggende materiaal. De conclusies van Significant zijn in die zin dus wetenschappelijk verantwoord.
Let wel: de korte periode en de gegevensbeperkingen waarmee Significant heeft moeten werken laten geen harde, definitieve conclusies toe. Het is dan ook geboden om over een langere periode de registraties te verzamelen en bovendien te toetsen aan andere typen gegevens (pv's / interviews), om te bezien of uitkomsten dezelfde orde van grootte hebben als de korte-termijn-resultaten. Dit is een van de opgaven voor de Lange-termijn Monitor.

Onderzoeksafdeling Rechtsbestel, Wetgeving, Internationale- en vreemdelingen-aangelegenheden (RWI)

Datum
5 oktober 2016

Ons kenmerk
2000381