

# Koppelmogelijkheden van ambulancedata met andere bronnen

R-2020-15

# SWOV



## Auteurs



Dr. L.T. Aarts



Dr. G.J. Wijlhuizen



Dr. F. Hermens



Drs. N.M. Bos

Ongevallen **voorkomen**  
Letsel **beperken**  
Levens **redden**

---

## Documentbeschrijving

Rapportnummer:	R-2020-15
Titel:	Koppelmogelijkheden van ambulancedata met andere bronnen
Auteur(s):	Dr. L.T. Aarts, dr. G.J. Wijlhuizen, dr. F. Hermens & drs. N.M. Bos
Projectleider:	Dr. L.T. Aarts
Projectnummer SWOV:	E20.13
Projectcode opdrachtgever:	31159349
Opdrachtgever:	Rijkswaterstaat

**Projectinhoud:** De politie- en ziekenhuisregistratie van verkeersongevallen- en -slachtoffergegevens is om uiteenlopende redenen niet compleet. Landelijke ambulancegegevens over verkeersongevallen kunnen een goede aanvulling zijn op deze registraties. SWOV is daarom gevraagd om uit te zoeken 1) hoe ambulancedata kunnen worden gekoppeld met andere gegevensbronnen en 2) wat er nodig is om het resultaat daarvan van toegevoegde waarde te laten zijn voor het verkeersveiligheidsbeleid. Dit rapport doet verslag van dat onderzoek en eindigt met een aantal aanbevelingen.

**Aantal pagina's:** 59  
**Fotografen:** Paul Voorham (omslag) – Peter de Graaff (portretten)  
**Uitgave:** SWOV, Den Haag, 2020

**De informatie in deze publicatie is openbaar.  
Overname is toegestaan met bronvermelding.**

**SWOV – Instituut voor Wetenschappelijk Onderzoek Verkeersveiligheid**

Bezuidenhoutseweg 62, 2594 AW Den Haag – Postbus 93113, 2509 AC Den Haag  
070 – 317 33 33 – [info@swov.nl](mailto:info@swov.nl) – [www.swov.nl](http://www.swov.nl)

 [@swov\\_nl](https://twitter.com/swov_nl) / [@swov](https://twitter.com/swov)  [linkedin.com/company/swov](https://www.linkedin.com/company/swov)

## Samenvatting

Verkeersongevallen- en slachtoffergegevens vormen een belangrijke basis voor verkeersveiligheidsbeleid. Twee belangrijke landelijk beschikbare bronnen van deze gegevens zijn het Bestand geRegistreerde Ongevallen in Nederland (BRON, voornamelijk gebaseerd op gegevens van de politie, en derhalve bekend als 'de politieregistratie') en de Landelijke Basisregistratie Ziekenhuiszorg (LBZ, 'de ziekenhuisregistratie'). Een koppeling van deze bestanden is ook de basis voor de schatting van het jaarlijks aantal ernstig verkeersgewonden door SWOV.

BRON en de LBZ zijn om uiteenlopende redenen niet compleet. Uit eerdere pilots is gebleken dat landelijke ambulancegegevens over verkeersongevallen een goede aanvulling kunnen zijn op deze en andere bestanden met ongevallen- en slachtoffergegevens. Hoewel de verrijkte bestanden nog steeds niet compleet zullen zijn, is de verwachting dat een succesvolle koppeling met BRON leidt tot een factor 3 meer registraties van verkeersongevallen en -slachtoffers. Daarnaast is de inschatting dat voor circa 80% van de in ziekenhuizen opgenomen verkeersslachtoffers ook ambulancegegevens beschikbaar zijn. Voor deze verkeersslachtoffers kunnen de ambulancedata een sleutel bieden voor meer informatie over het ongeval waarbij ze betrokken waren (bijvoorbeeld de locatie van het ongeval).

Het verwerken en ontsluiten van ambulancedata voor verkeersveiligheidsdoeleinden is dan ook opgenomen als maatregel in het *Landelijk Actieplan Verkeersveiligheid 2019-2021*.<sup>1</sup> Als voorbereiding hierop heeft Rijkswaterstaat aan SWOV gevraagd om uit te zoeken hoe ambulance-data kunnen worden gekoppeld met andere bronnen. De hoofdvragen van Rijkswaterstaat kunnen als volgt worden samengevat:

- Hoe kunnen ambulancedata worden gekoppeld met andere bronnen, in het bijzonder met BRON en de LBZ?
- Hoe zorgen we ervoor dat het resultaat van de koppeling van toegevoegde waarde is voor beleidsmakers?

## Werkwijze

Om de vragen van Rijkswaterstaat te beantwoorden, hebben we eerst een inventarisatie uitgevoerd van de informatiebehoefte van beleidsmakers en direct betrokkenen bij de ambulance-registraties (onder andere via online interviews en een vragenlijst in samenwerking met VeiligheidNL). Daarna hebben we de mogelijkheden en voorwaarden verkend om ambulancedata succesvol te koppelen met andere bronnen. Hiervoor hebben we onder meer gebruikgemaakt van de bij SWOV beschikbare kennis over koppeling van databestanden en de kenmerken van de verschillende databestanden. In deze inventarisatie is ook relevante documentatie betrokken, waaronder eerder uitgevoerd onderzoek en informatie over wetgeving en de implicaties daarvan op de koppeling van gegevens.



<sup>1</sup> Ministerie van Infrastructuur en Waterstaat (2018). *Landelijk Actieplan Verkeersveiligheid 2019-2021. Veilig van deur tot deur*. Ministerie van Infrastructuur en Waterstaat, Den Haag.

## Informatiebehoefte van beleidsmakers

Beleidsmakers maken voor hun beleid vooral gebruik van data uit BRON. Voor hen kunnen ambulancegegevens dan ook een belangrijke aanvulling zijn om BRON te verrijken. Het merendeel van de ondervraagde beleidsmakers en wegbeheerders geeft aan behoefte te hebben aan landelijke ontsluiting van een dergelijk verrijkt bestand, inclusief de nodige duidingen en achtergrondinformatie. Dit kan bijvoorbeeld worden vormgegeven met een landelijk openbaar dataportaal.

Daarnaast hebben beleidsmakers behoefte aan advies over het gebruik van de data. Dat zou kunnen door een adviesloket in te richten, bijvoorbeeld onder de vlag van het Kennisnetwerk SPV (Strategisch Plan Verkeersveiligheid).

Een deel van de beleidsmakers wil ook zelf analyses uitvoeren met de beschikbare data. Daarom hebben zij behoefte aan data op hoog detailniveau, met gegevens over het ongeval (datum, tijd, locatie, omstandigheden), maar ook over kenmerken van slachtoffers en betrokkenen (aantal, leeftijd, geslacht, nationaliteit, overtredingen zoals rijden onder invloed, vervoerswijze/tegenpartij, letselerst van slachtoffers). Dit soort gedetailleerde gegevens staan echter op gespannen voet met de privacywetgeving rond patiëntinformatie. Om de anonimiteit te waarborgen, zullen persoonsgegevens versleuteld moeten worden.

## Koppelmogelijkheden

De mogelijkheden om ambulancegegevens te koppelen aan die in BRON, de LBZ of andere databestanden, zijn ten eerste mede afhankelijk van de doelstelling: is het doel om een bestand zoals BRON te verrijken, of willen we een beter beeld krijgen van een specifiek probleem, zoals het jaarlijkse aantal ernstig verkeersgewonden? In dit onderzoek hebben we verkend wat nodig is om deze beide doelen te verwezenlijken.

Ten tweede is het van belang welke kenmerken binnen een bepaalde databron beschikbaar zijn: hoe gedetailleerder de data en hoe preciezer bepaalde gegevens van individuele slachtoffers of ongevallen overeenkomen tussen verschillende bestanden, hoe 'beter' gegevensbestanden aan elkaar gerelateerd kunnen worden. BRON bevat bijvoorbeeld momenteel geen unieke persoonskenmerken om betrokken personen te kunnen identificeren, zoals een burgerservicenummer (BSN). De ruwe politiedata die (deels) ten grondslag liggen aan BRON bevatten die wel. Daarmee zouden we dan een betrouwbaardere koppeling met ambulancegegevens kunnen maken dan wanneer we alleen zouden beschikken over minder unieke persoonsgegevens zoals leeftijd en geslacht.

De beschikbaarheid van bepaalde gegevens – alsmede het detailniveau ervan – is niet alleen afhankelijk van het feit of deze gegevens in een databestand voorkomen, maar ook van de vraag of de eigenaar van het bestand deze gegevens en het gevraagde detailniveau daarvan beschikbaar willen stellen, bijvoorbeeld voor onderzoek of gebruik door beleidsmakers. Hierbij is vooral de interpretatie van de privacywetgeving van belang: wat beschouwt de gegevenseigenaar als 'tot personen te herleiden' informatie?

Voor een daadwerkelijke koppeling tussen regels van verschillende bestanden is het vooral van belang dat slachtoffers *uniek onderscheidbaar* moeten zijn, maar *niet herleidbaar*. Dit is momenteel niet het geval in de ambulancedata die beschikbaar zijn gesteld als open data (Flevoland) of met wat meer detailniveau voor specifiek verkeersveiligheidsonderzoek. Unieke onderscheidbaarheid zonder herleidbaarheid kan worden bereikt door uniek onderscheidbare gegevens te versleutelen.



## Benodigde gegevens voor koppeling van ambulancedata met andere bronnen

In onderstaande tabel zijn de mogelijkheden samengevat van welke bronnen op basis van welke gegevens gekoppeld kunnen worden.

*Koppelmogelijkheden van verschillende bronnen en daarbij benodigde koppelgegevens. PC6/PC 4 staat voor respectievelijk 'postcode compleet' (PC6) of alleen de cijfercode daaruit (PC4).*

Koppelmogelijkheden	Combinatie met ambulancedata		
	BRON	Ruwe politiedata	LBZ
Koppeling op basis van BSN	Momenteel niet mogelijk vanwege het ontbreken van het BSN in BRON. Het BSN komt in de toekomst mogelijk weer wel in BRON beschikbaar; in dat geval is versleuteling nodig.	Mogelijk, wel toestemming van JenV nodig aan politie en verwerker voor levering van deze gegevens; versleuteling van BSN nodig; vooral onderzoekspotentie (o.a. verbetering BRON via andere maatschappelijke klassen en tekstherkenning); hoge koppelkwaliteit.  Vooralsnog vooral interessant voor onderzoeksdoeleinden en indirect verbetering van BRON.	Mogelijk, wel versleuteling van BSN nodig; hoge koppelkwaliteit.  Vooralsnog vooral interessant voor onderzoeksdoeleinden en verbetering van vaststelling van het aantal ernstig verkeersgewonden en kenmerken van het ongeval waarin ze betrokken waren.
Koppeling op basis van (persoons)kenmerken	Wordt reeds toegepast in koppeling met de LBZ; voor hoge koppelkwaliteit en nuttige resultaten zijn volgende gegevens nodig: <ul style="list-style-type: none"> <li>&gt; ongevalsdatum en -tijd;</li> <li>&gt; locatie van de ambulance (gps, of anders PC6 + straat+ huisnummer of desnoods PC4 + straat);</li> <li>&gt; leeftijd slachtoffer (geboortedatum of anders leeftijd + geboortedag of leeftijd + geboortemaand) en geslacht;</li> <li>&gt; vervoermiddel van slachtoffer en tegenpartij.</li> </ul> Deze variant is vooralsnog het meest aan te bevelen voor de verrijking van BRON.	Zie BRON.	Wordt reeds toegepast in koppeling met BRON; voor hoge koppelkwaliteit en nuttige resultaten zijn volgende gegevens nodig: <ul style="list-style-type: none"> <li>&gt; ongevalsdatum en -tijd</li> <li>&gt; BSN of leeftijd slachtoffer (geboortedatum of anders leeftijd + geboortedag of leeftijd + geboortemaand) en geslacht;</li> <li>&gt; vervoermiddel van slachtoffer en tegenpartij;</li> <li>&gt; vervoerd naar ziekenhuis;</li> <li>&gt; naam ziekenhuis.</li> </ul> Deze variant is vooralsnog aan te bevelen voor de verrijking van ernstig verkeersgewonden in het bestaande proces met BRON.
Combinatie op basis van open data (niet herleidbaar tot personen)	Geen koppeling mogelijk, alleen combinaties van overeenkomende clusters, vooral op basis van overeenkomst in ongevalslocatie.	Zie BRON.	Geen koppeling mogelijk wegens ontbreken persoonsgegevens.

## Koppelmethoden

Voor verrijking van BRON zijn deterministische koppelmethoden, bij voorkeur met afstandsfunctie aan te bevelen. Hierbij wordt gekeken naar zo veel mogelijk overeenkomst in kenmerken van slachtoffers en/of ongevallen om een koppeling vast te stellen tussen regels binnen bestanden.

Voor verrijking van ernstig verkeersgewonden wordt vooralsnog vastgehouden aan koppeling met afstandsfunctie. Op termijn is eventueel een probabilistische koppeltechniek interessant, waarbij gekoppeld wordt op basis van kansen van voorkomen van kenmerken. Deze mogelijkheid verkent SWOV nog nader. Een extra stap bij de vaststelling van werkelijke aantallen slachtoffers is de bijschattingsprocedure, waarbij wordt vastgesteld welk deel van slachtoffers of ongevallen in de gekoppelde bestanden nog ontbreekt. Deze bijschattingsprocedure is al een vast onderdeel bij de vaststelling van het aantal ernstig verkeersgewonden in Nederland.

## Aanbevelingen

Om ambulancegegevens te koppelen aan andere bronnen, is het in de eerste plaats van belang om te weten wat de doelen daarvan zijn (vooral BRON verrijken ten behoeve van wegbeheerders of ook een betere schatting maken van het aantal ernstig verkeersgewonden). In dit onderzoek hebben we verkend wat nodig is om deze beide doelen te verwezenlijken. We bevelen aan om hierbij ook de ruwe politieregistratie te betrekken die (deels) ten grondslag liggen aan BRON.

Voor een goede koppeling met BRON en/of de LBZ en als bouwstenen voor een voor beleid bruikbaar eindresultaat, zijn minimaal de volgende ambulancedata nodig:

- > ongevalsinformatie;
- > ongevalsdatum en -tijd;
- > locatie van de ambulance (gps, of anders PC6 + straat+ huisnummer of desnoods PC4 + straat);
- > slachtofferinformatie;
- > leeftijd slachtoffer (geboortedatum of anders leeftijd-geboortedag of leeftijd-geboortemaand) en geslacht;
- > vervoermiddel van slachtoffer en tegenpartij;
- > vervoerd naar ziekenhuis (voor koppeling met de LBZ);
- > naam ziekenhuis (voor koppeling met de LBZ).

De keuzes die hierbij gemaakt moeten worden, hebben consequenties voor de partijen die bij het proces betrokken zijn. In de eerste plaats moet met de ambulancediensten en de koepelorganisaties worden afgesproken welke gegevens beschikbaar kunnen worden gesteld, met welk doel en op welke wijze. Daarnaast moet worden onderzocht voor welke data er toestemming moet zijn om deze te mogen gebruiken voor verbetering van de verkeersveiligheid. Ook is het van belang om uit te zoeken of bij de versleuteling van persoonsgegevens een onafhankelijke derde partij moet worden betrokken, en wie deze 'Trusted Third Party' (TTP) dan zou moeten zijn.

Vanuit deze vragen bevelen we aan om de koppelmogelijkheden van ambulancedata met andere bronnen te verkennen via een proeflevering. Daaruit moet blijken welke extra details beschikbaar kunnen komen over ongevallen en slachtoffers die gebruikt kunnen worden voor verder onderzoek en beleid. Deze inzichten kunnen vervolgens ook worden gebruikt om stapsgewijs verder te werken aan een landelijk traject.

## Inhoud

<b>1</b>	<b>Inleiding</b>	<b>10</b>
<b>2</b>	<b>Informatiebehoefte van beleidsmakers</b>	<b>12</b>
2.1	Methode	12
2.1.1	Interviews en vragenlijst:	12
2.1.2	Aanvullende bronnen	13
2.2	Resultaten	13
2.2.1	Bevindingen uit eerdere inventarisaties	13
2.2.2	Resultaten uit de interviews	14
2.2.3	Resultaten uit de vragenlijst	15
2.3	Samenvattend: informatiebehoefte van beleidsmakers	16
<b>3</b>	<b>Combineren van ambulancedata met andere bronnen</b>	<b>17</b>
3.1	Gebruikte informatie en methoden	17
3.2	Bronnen en hun (combinatie)kenmerken	18
3.2.1	BRON	18
3.2.2	Ruwe politiegegevens	19
3.2.3	Ziekenhuisregistratie LBZ	20
3.2.4	Ambulancegegevens	20
3.3	Koppelscenario's	21
3.3.1	Koppelen en het doel van de koppeling van bronnen	21
3.3.2	Beschikbare data en wat dit betekent voor koppelmogelijkheden	23
3.3.3	Technieken om bestanden te koppelen en werkelijke omvang vast te stellen	28
3.4	Samenvattend: koppelmogelijkheden en vereisten vanuit de data	31
3.4.1	Verrijking van BRON	31
3.4.2	Koppeling met ruwe politiegegevens	32
3.4.3	Koppeling met de LBZ	32
<b>4</b>	<b>Aanbevelingen</b>	<b>35</b>
	<b>Literatuur</b>	<b>37</b>
<b>Bijlage A</b>	<b>Koppelvraag in de online vragenlijst</b>	<b>38</b>
<b>Bijlage B</b>	<b>Kenmerken per bron en koppelmogelijkheden</b>	<b>39</b>
<b>Bijlage C</b>	<b>Verwerking van BSN</b>	<b>45</b>
<b>Bijlage D</b>	<b>Vereisten bij de verwerking van patiëntgegevens</b>	<b>46</b>
<b>Bijlage E</b>	<b>Anonimisering van gegevens via een TTP</b>	<b>49</b>



<b>Bijlage F</b>	<b>Beschikbare data vanuit ambulanceregio's: een casus</b>	<b>51</b>
<b>Bijlage G</b>	<b>Schatting van overlap tussen registraties</b>	<b>53</b>
<b>Bijlage H</b>	<b>Toelichting op de koppelmogelijkheden van open data</b>	<b>55</b>

## 1 Inleiding

**Verkeersongevallen- en slachtoffergegevens zijn een belangrijke basis voor verkeersveiligheidsbeleid. Twee belangrijke landelijk beschikbare bronnen van deze gegevens zijn het Bestand geRegistreerde Ongevallen in Nederland (BRON, voornamelijk gebaseerd op gegevens van de politie, en derhalve bekend als ‘de politieregistratie’) en de Landelijke Basisregistratie Ziekenhuiszorg (LBZ, ‘de ziekenhuisregistratie’). Uit eerdere pilots is gebleken dat landelijke ambulancedata over verkeersongevallen een goede aanvulling kunnen zijn op deze bestanden.**

Het verwerken en ontsluiten van ambulancedata voor verkeersveiligheidsdoeleinden is opgenomen als maatregel (18) in het *Landelijk Actieplan Verkeersveiligheid 2019-2021* (Ministerie van Infrastructuur en Waterstaat, 2018). Als voorbereiding hierop heeft Rijkswaterstaat aan SWOV gevraagd om uit te zoeken hoe ambulancedata succesvol kunnen worden gekoppeld aan bestaande data over verkeersongevallen en -slachtoffers. Daarbij is het ook van belang te kijken naar de vraag wat dit betekent voor de kwaliteit van de data en de bruikbaarheid ervan voor beleidsmakers.

In dit rapport gaan we stap voor stap in op de vragen van Rijkswaterstaat, die we hier opsplitsen in twee delen.

### **Kwaliteitswensen van beleidsmakers**

1. Hoe zorgen we ervoor dat het eindproduct van gekoppelde data van toegevoegde waarde is voor beleidsmakers?
2. Welk detailniveau is daarbij gewenst?

### **Koppeling van ambulancedata met andere bronnen:**

3. Hoe is de koppeling tussen ambulancedata en andere bronnen (met name BRON en de LBZ) mogelijk?
4. Wat is het minimaal benodigde kwaliteits- en detailniveau van ambulancedata om koppeling met andere databronnen mogelijk te maken?
5. Hoe dragen we er zorg voor dat de vastgestelde koppelcriteria zo goed mogelijk worden meegenomen bij het verzamelen en ontsluitingen van ambulancedata als bron op zich ?
6. Wat is de rol van de verschillende partijen daarbij?

### **Leeswijzer**

De vragen van Rijkswaterstaat komen achtereenvolgens aan de orde in *Hoofdstuk 2* (informatiebehoefte van wegbeheerders) en *Hoofdstuk 3* (koppelmogelijkheden). We sluiten de notitie af met een aantal aanbevelingen (*Hoofdstuk 4*).

## Toelichting op terugkerende kernbegrippen in deze notitie:

Begrip	Betekenis
Combineren	<p>Verzamelbegrip, duidend op de activiteit waarbij bestanden aan elkaar gerelateerd worden.</p> <p>Doel: op basis hiervan kan meer of minder gedetailleerd de overlap tussen bestanden bepaald worden.</p>
Koppelen	<p>Een specifieke vorm van combineren, namelijk een-op-een combineren van regels in het ene bestand met regels in het andere bestand.</p> <p>Doel: gedetailleerd bepalen van de overlap tussen bestanden. In het eerste deel van de notitie wordt nog over 'koppelen' gesproken als ideaalbeeld; in het tweede deel van de notitie spreken we van koppelen als het daadwerkelijk om koppelen gaat, anders spreken we over combineren van bestanden.</p>
Verrijken	<p>Koppelen van twee of meer bestanden waarbij een nieuw gecombineerd gebruikersbestand ontstaat.</p> <p>Noot: een verrijkt bestand kan nog steeds onderregistraties bevatten en geeft niet zonder meer een beeld van werkelijke aantallen slachtoffers of ongevallen. Hiervoor is namelijk nog een stap extra nodig, waarbij een schatting wordt gegeven van wat in de betrokken bestanden ontbreekt. Dit noemen we de 'bijschatting'. Het verrijkte bestand plus de bijschatting geven een beeld van het werkelijke aantal slachtoffers.</p>
Ontdubbelen	<p>Een stap die zowel binnen een bestand als na koppelen van bestanden kan plaatsvinden.</p> <p>Doel: regels die op hetzelfde slachtoffer of ongeval betrekking hebben niet meer dan één keer te laten meetellen. Iedere regel vertegenwoordigt dan na ontdubbeling één slachtoffer of één ongeval.</p>

## 2 Informatiebehoefte van beleidsmakers

**Beleidsmakers willen een beter beeld krijgen van de werkelijke omvang van verkeersongevallen en -slachtoffers en van de bijbehorende kenmerken. De vraag daarbij is in hoeverre dat mogelijk is door ambulancegegevens te koppelen aan andere ongevallen en slachtofferregistraties, zoals BRON en de LBZ.**

Verschillende registraties kunnen elkaar enerzijds aanvullen (welke extra informatie geeft een andere bron?) en anderzijds overlappen. Wat betreft dit laatste punt kan koppeling zorgen voor ‘ontdubbeling’ van geregistreerde verkeersongevallen. Voordat we de koppelmogelijkheden zelf bespreken (Hoofdstuk 3), gaan we in de volgende paragrafen eerst in op het belang van een koppeling voor beleidsmakers.

### 2.1 Methode

We hebben de informatiebehoefte van wegbeheerders en andere beleidsmakers op drie manieren in kaart gebracht:

1. online interview;
2. een online vragenlijst;
3. aanvullende schriftelijke bronnen.

De interviews en vragenlijst zijn afgenomen in samenwerking met VeiligheidNL, dat parallel aan de SWOV-studie een eigen onderzoek deed naar de vraag hoe ambulancegegevens zijn te ontsluiten en hoe deze optimaal kunnen aansluiten bij de wensen van wegbeheerders.

#### 2.1.1 Interviews en vragenlijst:

SWOV en VeiligheidNL hebben begin juni 2020 vier online interviews (via een beeldverbinding) afgenomen met vertegenwoordigers van:

- Provincie Gelderland en Regionaal Orgaan Verkeersveiligheid Oost Nederland;
- Provincie Noord-Holland en Vervoerregio Amsterdam;
- Provincie Zuid-Holland en Metropoolregio Rotterdam Den Haag;
- Provincie Noord-Brabant.

De gesprekken werden gevoerd aan de hand van een door VeiligheidNL en SWOV opgestelde gespreksleidraad. In de gesprekken zijn meerdere onderwerpen aan de orde geweest die betrekking hebben op ambulancegegevens ten behoeve van verkeersveiligheid. Eén van de onderwerpen betrof de relevantie van het koppelen van ambulancegegevens met andere ongevallen- en slachtofferregistraties. De antwoorden op dit gespreksonderwerp worden hier weergegeven. Voor de bespreking van de andere onderwerpen die aan bod zijn geweest, verwijzen we naar VeiligheidNL (Nijman et al., 2020).

In de online vragenlijst is gevraagd hoe zinvol de betreffende wegbeheerder het zou vinden als er gekoppelde data beschikbaar zouden komen en waarom (zie *Bijlage A*). In totaal hebben 54 respondenten de vragenlijst ingevuld. Daarvan zijn er 40 afkomstig van gemeenten en 14 van provincies, vervoerregio's en regionale organen verkeersveiligheid (ROV's).

### 2.1.2 Aanvullende bronnen

Als bronnenmateriaal is voor dit onderdeel vooral geput uit het *Eindrapport kwaliteitsverbetering informatieketen verkeersongevallenregistratie* (Rijkswaterstaat, 2018). Dit rapport is gebaseerd op een onderzoek van Rijkswaterstaat met als titel: *Verbetering Informatieketen Verkeersongevallen*. De aanleiding voor dit onderzoek was de zorg over de afgenomen kwaliteit van de registratie van verkeersongevallen. Rijkswaterstaat heeft dit onderwerp mede opgepakt als beheerder van het BRON-bestand, dat wordt samengesteld op basis van de politieregistratie en door Rijkswaterstaat beschikbaar wordt gesteld aan beleidsmakers, onderzoekers en maatschappelijke organisaties (zie *Paragraaf 3.2.1*).

## 2.2 Resultaten

In deze paragraaf bespreken we achtereenvolgens de resultaten uit eerdere inventarisaties, de interviews en de vragenlijst.

### 2.2.1 Bevindingen uit eerdere inventarisaties

Uit het kwaliteitsverbeteringsrapport van Rijkswaterstaat (Rijkswaterstaat, 2018) komt een aantal aanbevelingen naar voren die relevant zijn voor de wenselijkheid van het gebruik van ambulancegegevens voor verkeersveiligheidsbeleid. Het gaat om de volgende aanbevelingen:

1. *“Maak ambulancedata toegankelijk voor verkeersveiligheidsbeleid. Ambulancedata vult de huidige ongevalsregistratie BRON aan met informatie over ongevallen zonder motorvoertuigen. Als de gegevens in de verdere toekomst gekoppeld kunnen worden aan BRON zou dat ook helpen om te voldoen aan de vraag van de Europese Commissie voor de vulling van de Europese ongevalsregistratie CARE.”*
2. *“Hanteer de als ‘need to have’ aangemerkte informatiebehoefte per doelgroep [wegbeheerders en andere overheden betrokken bij verkeersveiligheid] als vertrekpunt voor de te leveren gegevens in het nieuwe productbeleid.”*  
(Rijkswaterstaat, 2018; p. 80)

Ten aanzien van punt 2 is een aantal kenmerken niet op voorhand nodig op het niveau van individuele ongevallen of slachtoffers, maar wel als statistiek op geaggregeerd niveau. Omdat de uitlevering van detailinformatie over ongevallen en slachtoffers aan beleidsmakers niet zonder meer mogelijk is (vanwege de Algemene verordening gegevensbescherming, AVG), zouden de volgende statistieken per gemeente geaggregeerd geleverd moeten worden om verder in de informatiebehoefte voor verkeersveiligheidsbeleid te voorzien:

- Aantal slachtoffers naar letselnst, leeftijdsklasse (liefst zo fijnmazig mogelijk) en geslacht. Deze zijn nodig voor het bepalen van trends en algemene statistieken.
- Aantal ongevallen met rijden onder invloed van alcohol. En in de toekomst als het mogelijk is ook vereist voor drugs- en medicijngebruik en gebruik van apparatuur bij ongevallen. Inmiddels is een wetgevingstraject gestart om deze gegevens ook in meer detail aan beleidsmakers te kunnen leveren.

Wegbeheerders, decentrale overheden en vervoerregio's noemen de volgende kenmerken als noodzakelijke informatiebehoefte. De met een \* aangegeven kenmerken zijn in potentie ook aanwezig in de Basisset Ambulancezorg (BSA).<sup>2</sup>

- > Locatie-informatie\*
- > Wegkenmerken
- > Datum en tijd\*
- > Letselernst
- > Aantal slachtoffers\*
- > Slachtofferkenmerken\*
- > Gegevens per betrokkene / partij
- > Vervoerswijze en objecttype\*
- > Leeftijd bestuurder\*<sup>3</sup>
- > Geslacht bestuurder\*<sup>3</sup>
- > Rijden onder invloed
- > Nationaliteit
- > Ongevalsekenmerken
- > Omstandigheden

Uit de weergegeven lijst blijkt dat een aantal noodzakelijk geachte kenmerken nog algemeen is geformuleerd (bijvoorbeeld: ongevalskenmerken). De publicatie biedt op dat punt dus geen verdere duidelijkheid over welke informatie over ongevalskenmerken precies gewenst is. Uit de interviews zal echter blijken dat beleidsmakers zo veel mogelijk willen weten omdat dit kan helpen het ongevalsproces en het daarop aansluitende beleid te formuleren.

## 2.2.2 Resultaten uit de interviews

Uit de interviews zijn de volgende resultaten gekomen die van belang zijn als het gaat om koppeling van ambulanceregistraties met andere registraties zoals BRON:

- > De ongevallenregistratie BRON (veelal geraadpleegd via applicaties van bureau VIA) wordt gewoonlijk (al meerdere jaren) gebruikt als basis voor het verkeersveiligheidsbeleid. Door ambulancegegevens te koppelen aan BRON-gegevens, wordt aangesloten bij gegevens die bekend zijn bij beleidsmakers (zie ook *Paragraaf 3.2.1*).
- > In drie van de vier interviews kwam naar voren dat er behoefte is aan (meer) gedetailleerde gegevens en het zelf kunnen analyseren daarvan. Genoemd werd onder meer dat er wat betreft fietsongevallen onvoldoende gegevens beschikbaar zijn (aantal, locatie, tijdstip, indicatie ernst, manoeuvre, tegenpartij). Als (combinaties van) verschillende registraties aanvullende informatie kunnen geven, bijvoorbeeld over fietsongevallen, dan is dat van belang voor de beleidsmakers. Dat maakt dat detailgegevens van ongevallen- en slachtofferkenmerken wenselijk zijn.
- > Het koppelen van ambulancegegevens aan bijvoorbeeld BRON werd in de drie interviews van belang geacht, met name nadat werd aangegeven wat een koppeling zou kunnen bewerkstelligen (zoals: ontdebellen, verrijken van BRON waarbij informatie uit ambulanceritten zou worden toegevoegd).
- > In één van de interviews werd aangegeven dat er geen behoefte bestaat aan een uitbereiding met informatie uit 'nieuwe' registraties zoals die van ambulancediensten. De reden die daarvoor werd aangevoerd, is dat het te onduidelijk is of deze informatie werkelijk nieuwe inzichten zou opleveren, terwijl de complexiteit van de gegevens zou toenemen vanwege de verschillende eigenschappen van de registraties.
- > Verschillende geïnterviewden gaven aan dat er op regionaal niveau afzonderlijke initiatieven zijn gestart om gegevens over verkeersongevallen te analyseren en te duiden (mede ingegeven



- 2 Of deze data daadwerkelijk aanwezig zijn, hangt af van de wijze van registratie van individuele ambulancediensten. Door alle ambulancediensten wordt de BSA wel als centrale basis benut, maar niet alle diensten noteren alles in de voorgeselecteerde variabelen. Ook invoer in vrije tekstvelden komt voor.
- 3 Alleen beschikbaar in de BSA indien er ambulance-inzet voor deze betrokkene(n) is geweest.



door het Strategisch Plan Verkeersveiligheid 2030 (Ministerie van Infrastructuur en Waterstaat et al., 2018). Hiervoor worden of zijn eigen 'dashboards' verkeersveiligheid in ontwikkeling of reeds beschikbaar. Niettemin gaven diverse geïnterviewden ook aan een belangrijk voordeel te zien in een landelijk gegevensbestand met daarin de nodige koppelingen van registraties, mede gezien de complexiteit van de verschillende registraties. Dit gegevensbestand zou toegankelijk moeten zijn voor beleidsmakers. De belangrijkste reden daarvoor is dat er dan voor alle lokale/regionale/provinciale/landelijke organisaties een eenduidige informatiebron kan worden geraadpleegd en deskundigheid daarover kan worden gebundeld. Dat is van belang omdat capaciteit voor verkeersveiligheidsbeleid op met name gemeentelijk niveau beperkt is.

### 2.2.3 Resultaten uit de vragenlijst

Vrijwel alle respondenten van de online vragenlijst (N=52) gaven aan dat ze het belang inzien van het koppelen van politie- en ambulanceregistraties van verkeersongevallen. Eén respondent gaf aan het niet te weten en één respondent heeft de vraag niet beantwoord. De respondenten konden in een tekstveld het antwoord toelichten. De daar gegeven toelichtingen kunnen in verschillende typen worden onderscheiden (Tabel 1).

Tabel 1. Respons van beleidsmakers op de online vragenlijst.

Type antwoord (meerdere antwoorden mogelijk per respondent)	Aantal keer genoemd
1 Beter totaalbeeld van de verschillende bronnen; wellicht in centraal systeem beheerd	24
2 Beter één (gekoppelde) gegevensbron gebruiken dan twee afzonderlijk; voorkomt dubbelstellingen en discussie	16
3 Meer detailinformatie ook voor diepgaander (risico)analyse en locaties ten behoeve van beleid	13
4 Zo mogelijk aanvullen met verzekeringsgegevens	2

De bovenstaande antwoorden laten zien dat er voordeel wordt gezien in een samengevoegd en gekoppeld (ontdubbeld) totaalbeeld. Daarnaast wordt het als een voordeel gezien dat het bestand centraal wordt beheerd, vooral om verwarring en discussie te voorkomen. Daarbij wordt ook aangegeven dat het verkeersveiligheidsbeleid kan worden versterkt als beleidsmakers door de koppeling kunnen beschikken over meer detailinformatie. Deze informatie kan beleidsmakers meer zicht bieden op oorzaken van ongevallen en zo betere aanknopingspunten bieden voor maatregelen.

De respondenten is daarnaast gevraagd hoe frequent ze een geactualiseerd beeld zouden willen hebben van de gekoppelde gegevens. In totaal 21 respondenten wilden dat bij voorkeur maandelijks, 19 per kwartaal en 10 respondenten één keer per jaar. Drie respondenten gaven aan dat ze geen belangstelling hebben voor deze gegevens. Formele registraties die zicht geven op verkeersongevallen en/of -slachtoffers, zoals BRON en de LBZ, worden momenteel één keer per jaar uitgeleverd.

## 2.3 Samenvattend: informatiebehoefte van beleidsmakers

Uit de interviews en vragenlijst is veelal naar voren gekomen dat er behoefte bestaat aan zo veel mogelijk gedetailleerde gegevens. Het gaat daarbij om informatie over:

- Het ongeval: datum, tijd, locatie, omstandigheden.
- Slachtoffers -en andere betrokkenen: aantal, leeftijd, geslacht, nationaliteit, overtredingen zoals rijden onder invloed, vervoerswijze/tegenpartij, letselerst van slachtoffers.

Hierbij is een zo groot mogelijk detailniveau gewenst. Diverse beleidsmakers geven daarvoor als reden aan graag zelf analyses uit te voeren met deze gegevens, om zo betere handvatten te hebben voor beleidskeuzen die passen bij de lokale en actuele verkeersveiligheidsproblemen.

Deze behoefte vertaalt zich momenteel in verschillende lokale en regionale initiatieven om gegevens over verkeersongevallen te kunnen analyseren en te duiden. Dit leidt onder meer tot de ontwikkeling van eigen 'dashboards' verkeersveiligheid. Niettemin wordt er over het algemeen een belangrijk voordeel gezien in een landelijk en voor beleidsmakers vrij toegankelijk gegevensbestand met daarin de nodige koppelingen van registraties. De belangrijkste reden daarvoor is dat er dan voor alle lokale/regionale/provinciale/landelijke organisaties een eenduidige informatiebron kan worden geraadpleegd en – gezien de complexiteit van de gegevens – deskundigheid daarover kan worden gebundeld. Deze bevindingen sluiten aan bij de aanbevelingen uit het kwaliteitsverbeteringsrapport van Rijkswaterstaat (Rijkswaterstaat, 2018). Hieraan zit ook een Europese dimensie. Zo zou ook de ongevallenregistratie in het EU-databestand CARE van een dergelijke koppeling kunnen profiteren, omdat voor CARE nu alleen gebruik wordt gemaakt van BRON voor wat betreft de Nederlandse situatie. Verbeteringen of verrijking van BRON zou dan ook een verbetering van de in EU-verband gebruikte data betekenen.

Ook kwam uit de interviews naar voren dat het niet voor iedereen voldoende duidelijk is of de ambulancegegevens werkelijk nieuwe inzichten zouden opleveren, terwijl de complexiteit van de gegevens zou toenemen vanwege de verschillende eigenschappen van de registraties. Mogelijk kunnen deze bezwaren worden weggenomen door helder inzicht te geven in het te verwachten verbeterpotentieel als ambulancegegevens aan met name BRON gekoppeld gaan worden. Wat komt er door toevoeging van ambulancegegevens en na ontdebelling van gegevens in potentie aan slachtoffers extra in beeld?<sup>4</sup> De kennisverspreiding hierover zou bijvoorbeeld kunnen worden opgepakt als onderdeel van de Task Force Verkeersveiligheidsdata, waarin wordt gewerkt aan een actueel overzicht van (de mogelijkheden van) informatiebronnen over ongevallen en slachtoffers.

Daarnaast is het van belang om helder te maken wat bij een daadwerkelijke koppeling van bestanden de resterende omissies zijn van het gekoppelde bestand. Doordat bij een gekoppeld bestand minder ongevallen of slachtoffers ontbreken, is het gemakkelijker om de totale omvang van ongevallen of slachtoffers te beschrijven in plaats van binnen de afzonderlijke bestanden te benoemen wat daar wel en niet in zit. Dit zou de complexiteit van de verschillende bestanden minder groot maken.

In het algemeen lijkt er dus een brede belangstelling te bestaan bij beleidsmakers voor het koppelen van ambulancegegevens met in ieder geval BRON. Een aantal respondenten geeft er blijk van zich bewust te zijn van de complexiteit van informatie uit verschillende registraties. Zij kiezen om die reden voor een samengevoegd bestand dat centraal (landelijk) beschikbaar komt en waarbij deskundigen zorgdragen voor advies en toelichting over de achtergronden van de gegevens. Een deel van de beleidsmakers heeft belangstelling om zelf de gegevens over de kenmerken van ongevallen te kunnen analyseren.



4 Wat die verbeterpotentie is, hangt mede af van de mogelijkheden om ambulancedata op detailniveau te kunnen verwerken en ook de toekomstige kwaliteit van BRON via andere ontwikkelingen dan verrijking met ambulancegegevens.

## 3 Combineren van ambulancedata met andere bronnen

In dit hoofdstuk staan we stil bij de wijze waarop ambulancedata kunnen worden gecombineerd met andere bronnen met informatie over verkeersongevallen en -slachtoffers. Hierbij kijken we ook naar de vraag wat de vereiste kwaliteits- en detailniveaus zijn om daadwerkelijk een koppeling tussen regels in bestanden mogelijk te maken.

We richten ons primair op een combinatie van ambulancegegevens met BRON (voornamelijk gebaseerd op de politieregistratie) en de LBZ (de ziekenhuisregistratie). Uiteindelijk staan we ook stil bij de vraag hoe de benoemde koppelcriteria zo goed mogelijk kunnen worden meegenomen in het vervolgproces van ontsluiting van ambulancegegevens en wat dit betekent voor de rol van verschillende partijen daarbij.

### 3.1 Gebruikte informatie en methoden

Om de vragen van Rijkswaterstaat (zie *Hoofdstuk 1*) te beantwoorden, hebben we onder meer gebruikgemaakt van de bij SWOV beschikbare kennis over koppeling van databestanden en de kenmerken van de verschillende databestanden. Dit betreft vooral de bestanden BRON, de achterliggende ruwe data van de politie en de LBZ.

SWOV heeft veel ervaring met het koppelen van BRON en de LBZ voor de jaarlijkse vaststelling van het aantal ernstig verkeersgewonden (zie bijvoorbeeld Reurings & Bos, 2009; Bos et al., 2019), en recentelijk ook met een koppeling van ruwe politiegegevens en de LBZ. Dit gebeurt de laatste jaren in de beveiligde microdata-omgeving van het CBS<sup>5</sup>, waardoor ook de ervaring met de procesgang aldaar in deze notitie is ingebracht.

Daarnaast hebben we voor dit rapport geput uit gesprekken met de volgende partijen:

- Ambulance Zorg Nederland (AZN; reeds in een eerdere fase uitgevoerd): dit heeft ons vooral inzicht gegeven in de vraag welke gegevens mogelijk in de Basisset Ambulancezorg (BSA) beschikbaar zijn (zie Schepers et al., 2018)<sup>6</sup>.
- Gezondheidsdienst Flevoland: hiermee is schriftelijk gecorrespondeerd over mogelijkheden en overwegingen om ambulancedata te kunnen leveren. Daarnaast is gekeken naar de ambulancedata die GGD Flevoland als open data reeds heeft gepubliceerd<sup>7</sup>. Ook is gebruikgemaakt van de kennis die SWOV heeft opgedaan met speciaal uitgeleverde onderzoeksdata van GGD Flevoland (zie onder andere Wijlhuizen & Bos, 2020; te verschijnen).



<sup>5</sup> <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen>

<sup>6</sup> Of deze data daadwerkelijk aanwezig zijn, hangt af van de wijze van registratie van individuele ambulancediensten. Door alle ambulancediensten wordt de BSA wel als centrale basis benut, maar niet alle diensten noteren alles in de voorgeselecteerde variabelen. Ook invoer in vrije tekstvelden komt voor.

<sup>7</sup> <https://www.eengezonderflevoland.nl/verkeersongevallen-flevoland-almere-lelystad> of <https://data.overheid.nl/dataset/verkeersongevallen-met-ambulancezorg-in-flevoland>; de ambulancedata van de ambulancedienst (RAV) Flevoland worden door de GGD Flevoland verwerkt.

- ZorgTTP (op 24 juni 2020): dit is een partij die een rol kan spelen bij versleuteling van gegevensbronnen. Hierdoor kunnen veilige en toch unieke koppelsleutels worden ontwikkeld.

## 3.2 Bronnen en hun (combinatie)kenmerken

In deze notitie richten we ons vooral op een combinatie van de ambulanceregistraties met BRON, aanvullende mogelijkheden via de ruwe politiedata en de LBZ. In deze paragraaf beschrijven we de meerwaarde van de verschillende bronnen. Ook geven we aan wat nog mogelijke barrières zijn om deze meerwaarde te kunnen benutten. In *Bijlage B* is een overzicht beschikbaar van de gegevens in de verschillende bronnen en op welke kenmerken gecombineerd zou kunnen worden.

### 3.2.1 BRON

Het Bestand geRegistreerde Ongevallen Nederland (BRON) is – zoals ook uit de interviews en documentatie blijkt – de informatiebron die beleidsmakers het meest gebruiken om een beeld van de verkeersveiligheid te krijgen. Als het gaat om ongevalskenmerken, is BRON op dit moment het meest uitgebreide bestand en zal dat naar verwachting ook blijven. BRON is hoofdzakelijk gebaseerd op de verkeersongevallenregistratie van de politie en wordt samengesteld en beheerd door Rijkswaterstaat. Rijkswaterstaat verrijkt deze gegevens met voertuigkenmerken van de RDW en lokaliseert de ongevallenlocaties op het Nationaal Wegenbestand (NWB). In gevallen waar de informatie in de vaste variabelen ontoereikend is, worden verbeterlagen uitgevoerd door onder meer uit de beschrijvende teksten van de politie bijvoorbeeld de exacte ongevalslocaties en vervoerswijze van slachtoffers vast te stellen.

Voordelen van BRON:

- Breed geaccepteerd in gebruik door wegbeheerders.
- Als open data beschikbaar<sup>8</sup>, al zijn daar de laatste jaren op grond van de AVG beperkingen in het detailniveau van de gegevens gekomen<sup>9</sup>.
- Biedt momenteel de meest uitgebreide set aan ongevalskenmerken, waaronder de locatie van ongevallen.

Nadelen van BRON:

- BRON is niet compleet. Met name een groot deel van de fietsslachtoffers die gewond raken ontbreekt (Reurings & Bos, 2009; zie ook het volgende punt voor meer inhoudelijke redenen hiervoor). Zo wordt het aandeel ernstig gewonde fietsslachtoffers dat we in BRON terugvinden, geschat op circa 10% van het totaal aantal ernstig gewonde fietsslachtoffers in ongevallen zonder motorvoertuig. Fietsslachtoffers in ongevallen met motorvoertuigen zijn voor circa 50% in BRON geregistreerd.
- Voor wat betreft de politieregistratie is BRON vooralsnog alleen gebaseerd op registraties van de politie die direct terug te voeren zijn op verkeersongevallen (de zogenoemde ‘maatschappelijke klassen’ D10 (verkeersongeval met uitsluitend materiële schade), D11 (verkeersongeval met letsel) en D12 (verkeersongeval met dodelijke afloop) die de politie in haar registratie hanteert; ongevallen kunnen ook onder andere maatschappelijke klassen worden geregistreerd). Zo weten we uit diepteonderzoek naar oudere fietsers (zie Tabel 7.6 in Davidse et al., 2014)) dat een deel van deze ongevallen werd geregistreerd onder met name E30: ongeval/onwel persoon (circa 36%). Ook voor bepaalde andere verkeersongevallen is bekend dat deze nogal eens onder andere maatschappelijke klassen in de ongevallenregistratie terechtkomt (bijvoorbeeld L6, spoorwegongevallen<sup>10</sup>).



8 Zie: <https://www.rijkswaterstaat.nl/apps/geoservices/geodata/dmc/bron/>

9 Geen slachtoffertabellen op individueel niveau meer beschikbaar. Zie ook Rijkswaterstaat (2018).

10 Persoonlijke communicatie Egbert Jan van Hasselt (politie), 28 maart 2019.

- BRON bevat momenteel geen unieke persoonsidentificatiegegevens zoals het burgerservice-nummer (BSN), omdat dergelijke gegevens alleen aan bepaalde instanties zijn voorbehouden om te verwerken en alleen indien dit voor dat werk noodzakelijk is<sup>11</sup>. Dit maakt een koppelproces bewerklijker dan wanneer deze gegevens wel aanwezig zouden zijn.

### 3.2.2 Ruwe politiegegevens

De ruwe vorm van de politieregistratie betreft een op onderdelen ander bestand dan BRON. Een deel van deze registratie is wel de basis voor BRON, maar kent niet de koppel- en verbeterlagen die Rijkswaterstaat voor BRON uitvoert. Anderzijds is het een uitgebreider bestand dat meer ongevallen en kenmerken omvat die onder voorwaarden kunnen worden benut (zie opmerkingen over onder andere E30 in de vorige subparagraaf).

Voordelen van ruwe politiegegevens:

- Meer potentiële verkeersongevallen beschikbaar dan alleen de ongevallen die bij de registratie als 'verkeersongeval' zijn aangemerkt (respectievelijk maatschappelijke klasse D10, D11 en D12). Relevante uitbreidingen zijn bijvoorbeeld D13 (doorrijden na ongeval), D72 (overige verkeerszaken), E30 (ongeval/onwel persoon) L6 (spoorwegongevallen)<sup>12</sup>.
- Meer gegevens beschikbaar, waaronder open tekstvelden en unieke identificatiekenmerken van betrokken personen (BSN) op basis waarvan betrokkenen in verschillende bestanden (met name de LBZ en ambulanceregistraties) exact aan elkaar gekoppeld kunnen worden.

Nadelen van ruwe politiegegevens:

- Ruwe politiegegevens zijn niet zonder meer beschikbaar. Hiervoor is toestemming nodig van het ministerie van Justitie en Veiligheid (JenV). De ervaring leert dat zij de volgende voorwaarden stellen aan politie (voor levering van de gegevens) en verwerkers (voor gebruik van de gegevens): a) resultaten zijn niet tot personen herleidbaar, b) het onderzoek dient het algemeen belang, c) de organisatie van de politie wordt niet onnodig belast, d) het onderzoek kan niet zonder de betrokken gegevens worden uitgevoerd en e) de persoonlijke levenssfeer van de betreffende personen wordt niet onevenredig geschaad<sup>13</sup>.
- Uitlevering van ruwe politiegegevens met bijzondere persoonsgegevens zoals het BSN is aan strenge eisen gebonden, net als vrije tekstvelden waarin gevoelige informatie kan staan. Organisaties mogen deze informatie niet zomaar verwerken (zie *Bijlage C*). Om indirect toch van deze gegevens gebruik te kunnen maken, dient de informatie ten minste versleuteld te worden via een onafhankelijke derde partij, een zogeheten Trusted Third Party zoals ZorgTTP of het CBS. Meer informatie over hoe dit in zijn werk gaat is te vinden in *Bijlage E*.
- Bij het benutten van ruwe politiegegevens die meer bevatten dan alleen de maatschappelijke klassen D10, D11 en D12, dient extra aandacht te worden besteed aan de selectie van verkeersongevallen, hetzij via opsporing van verkeerskenmerken in de vrije tekstvelden, hetzij via kenmerken uit andere bestanden die aan betrokken personen gekoppeld kunnen worden<sup>14</sup>.



11 Zie: <https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/identificatie/burgerservicenummer-bsn> (geraadpleegd 9 juli 2020).

12 Deze informatie biedt op termijn potentie om BRON te verbeteren. SWOV is voornemens om dit de komende jaren verder te gaan onderzoeken (concept-onderzoeksprogramma SWOV 2021. SWOV, Den Haag [in voorbereiding]).

13 Ervaringen SWOV op basis van justitiële beschikking van dergelijke data voor onderzoeksdoeleinden, 15 oktober 2018.

14 Op het moment dat personen in verschillende bestanden uniek aan elkaar gekoppeld kunnen worden (= het gaat met zekerheid om dezelfde persoon), is het nog niet zonder meer zeker dat deze persoon bij een verkeersongeval betrokken was als dat niet uit bepaalde kenmerken van de gekoppelde bestanden blijkt. Een persoon die zowel door de politie in een ongeval wordt geregistreerd en daarnaast ook door het ziekenhuis, kan bijvoorbeeld ook getuige zijn geweest bij een ongeval en toevallig enkele dagen later voor een niet-verkeersgerelateerde ingreep in het ziekenhuis zijn opgenomen. Uit bijvoorbeeld de betrokkenheid van voertuigen kan blijken of de betreffende registratieregeling een verkeersongeval betrof en of de betrokken persoon als slachtoffer bij het ongeval gewond raakte. Daarvoor is ook informatie nodig over de relatie tussen betrokkene en betrokken voertuigen.

### 3.2.3 Ziekenhuisregistratie LBZ

De Landelijke Basisregistratie Ziekenhuiszorg (LBZ, 'de ziekenhuisregistratie') dekt in principe 100% van alle behandelde patiënten die uit het ziekenhuis worden ontslagen. Dat wil zeggen dat bij ontslag van iedere patiënt in een ziekenhuis in Nederland, een registratie wordt opgemaakt die uiteindelijk in de LBZ terecht komt. Verbeterslagen en doorlevering van het bestand voor onderzoeksdoeleinden, gaat via Dutch Hospital Data (DHD).

Voordelen van de LBZ:

- De LBZ is vooralsnog het enige bestand waaruit betrouwbaar de ernst, aard en lichaamslocatie van het opgelopen letsel afgeleid kan worden. Dit is bijvoorbeeld van belang voor de relatie met beleidsdoelstellingen in termen van doden en ernstig verkeersgewonden.
- De LBZ bevat ook verkeersslachtoffers die op eigen gelegenheid naar het ziekenhuis komen en waar geen andere hulpdiensten aan te pas zijn gekomen. Dit zou naar schatting circa 10% zijn van alle verkeersslachtoffers die naar het ziekenhuis gaan (op basis van LIS 2009-2012; zie Wijlhuizen & Bos, te verschijnen). Met name veel fietserslachtoffers zonder betrokkenheid van gemotoriseerd verkeer vinden we zo terug.
- DHD heeft de mogelijkheid om ook unieke identificatiegegevens van betrokken personen (zoals het BSN) te leveren, overigens ook weer onder de eerder genoemde strikte voorwaarden.

Nadelen van de LBZ:

- De ziekenhuisregistratie bevat niet de locatie waar het ongeval heeft plaatsgevonden, omdat ziekenhuizen er geen belang bij hebben om dit voor hun processen te achterhalen en vast te leggen.
- Ongevalsekenmerken zijn slechts in zeer beperkte mate aanwezig en weinig gedetailleerd geregistreerd in de zogenoemde 'externe oorzaak' die is vervat in codes (zie bijvoorbeeld Bos et al., 2019)<sup>15</sup>. Deze codes bevatten een beperkte set vervoermiddelen, tegenpartijen en maken onderscheid naar bestuurder en passagier mogelijk.
- De LBZ wordt alleen uitgeleverd aan instanties waarmee de ziekenhuizen een overeenkomst hebben ten aanzien van gebruik en veilige beschikbaarstelling voor onderzoek. Dergelijke afspraken bestaan momenteel bijvoorbeeld met het CBS. Ook met organisaties als ZorgTTP wordt al samengewerkt.
- Bij benutting van unieke identificatiegegevens van betrokken personen (via het BSN), is versleuteling vereist (zie de informatie bij ruwe politiegelgegevens).

### 3.2.4 Ambulancegegevens

Voor verkeersongevallen waarvoor een ambulance is opgeroepen, is in principe een registratie bij de ambulanceorganisatie terug te vinden. Uit onderzoek blijkt dat circa 30% van de verkeersslachtoffers die naar het ziekenhuis of spoedeisende hulp gaan, met een ambulance binnenkomt (zie Wijlhuizen & Bos, te verschijnen). Dit betreft ook veel lichte letsels en letsels die op de spoedeisende hulp behandeld kunnen worden. Van de patiënten die daadwerkelijk worden opgenomen in een ziekenhuis, komt 80% met een ambulance (op basis van LIS 2009-2012; zie Wijlhuizen & Bos, te verschijnen). Omdat ambulances een eigen rol vervullen in de keten van ongeval-behandeling-afhandeling, bieden de ambulanceregistraties op specifieke onderdelen meerwaarde ten opzichte van de bronnen die al gebruikt worden en hiervóór benoemd zijn.

Voordelen van ambulancegegevens:

- Uit eerder onderzoek met ambulancegegevens (SWOV, niet gepubliceerd) is gebleken dat er circa 70.000 verkeersgerelateerde ambulanceritten per jaar plaatsvonden, tegen circa 19.000 letselongevallen in BRON (een factor 3 à 4 meer). In relatie tot de LBZ is de verwachting dat circa 80% van de verkeersslachtoffers ook in de ambulanceregistratie zullen voorkomen (zie



<sup>15</sup> Alle V-codes, maar ook enkele W,X, en Y-codes zijn van belang om verkeersongevallen te detecteren.



Wijlhuizen & Bos, te verschijnen). Deze cijfers kunnen worden benut om van een groot deel van de ernstig verkeersgewonden te achterhalen op welke locatie zij in een ongeval betrokken zijn geraakt. Met name voor de vele fietsslachtoffers in ongevallen zonder gemotoriseerd verkeer, zou dit waardevolle aanvullende inzichten bieden. Vooral deze groep is ondervertegenwoordigd in BRON.

- Daarnaast bieden ambulancedata ook mogelijkheden om meer kennis op te doen over de relatie tussen ongevals- en opnamelocaties en tijdsperiode tussen ongevals- en opnamelocatie. Dergelijke kennis kan leiden tot verbeteringen van de koppeling tussen BRON en de LBZ en daarmee van de vaststelling van het aantal ernstig verkeersgewonden.

Nadelen van ambulancegegevens:

- Registraties van ambulances zijn niet zonder meer beschikbaar voor beleidsmakers of onderzoekers. Dit heeft te maken met het feit dat ze (bijzondere) persoonsgegevens kunnen bevatten, zeker als voor het identificeren van verkeersongevallen in de bestanden vrije tekstvelden nodig zijn. Dit stelt voor gebruik door andere dan direct betrokken verwerkers extra eisen aan de voorbewerkingen van deze registratie, vooral ook om de vereisten uit de Wet op de geneeskundige behandelovereenkomst (WGBO) na te komen, die verder gaan dan de Algemene verordening gegevensbescherming (AVG, zie *Bijlage C*).
- Met afzonderlijke ambulancevoorzieningen moeten afspraken worden gemaakt over wat zij voor uitlevering en combinatie van gegevensbestanden juridisch acceptabel vinden. Er blijken momenteel nog verschillen tussen ambulanceorganisaties in uitwerking van hun visie op wat 'tot personen herleidbare informatie' is en wat dat betekent voor doorlevering of zelfs publicatie van bepaalde maten van geaggregeerde data. Dit blijkt bijvoorbeeld uit wat verschillende ambulancediensten aan data aan derden uitleveren: van GGD Flevoland die geaggregeerde ambulancedata als open data beschikbaar heeft gesteld, verschillende ambulanceregio's die voor onderzoeksdoeleinden iets meer gedetailleerdere gegevensbronnen beschikbaar stellen of stelden aan bijvoorbeeld SWOV en VeiligheidNL, tot een enkele ambulancedienst die daaraan tot nu toe nog helemaal niet wilde meewerken vanuit het bezwaar dat dit te veel inbreuk zou betekenen op de onthulling van persoonsgegevens.

### 3.3 Koppelscenario's

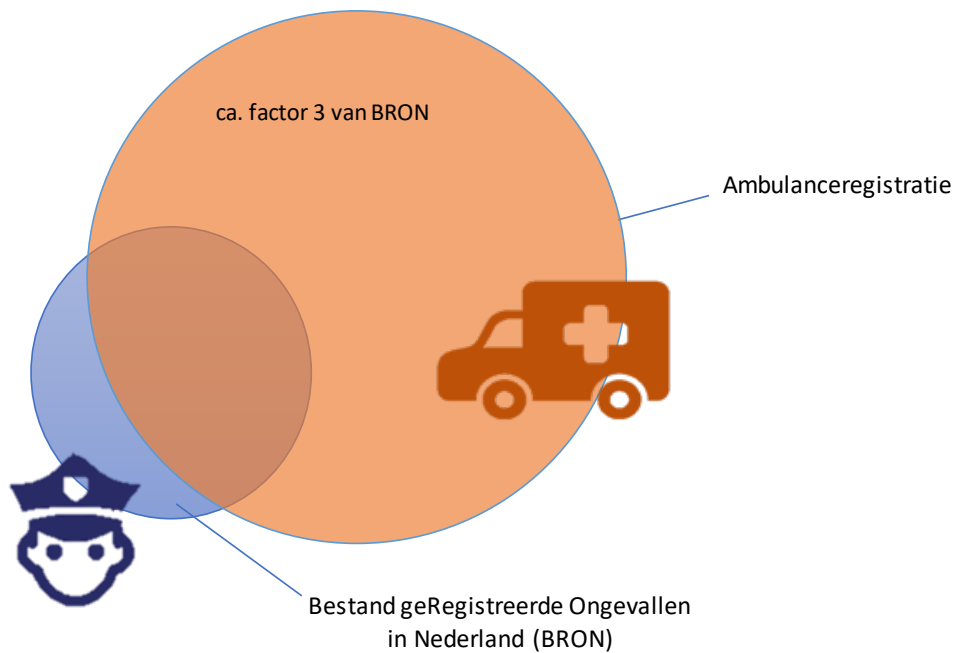
Er zijn verschillende mogelijkheden om bronnen met elkaar te combineren of te koppelen en zo een completer of beter beeld van de werkelijkheid te verkrijgen. Enerzijds spelen daarbij verschillende doelen van de combinatie of koppeling een rol, anderzijds spelen de beschikbare data en hun kenmerken een rol. Dit bepaalt in belangrijke mate wat mogelijk is. Ten derde zijn er technisch verschillende mogelijkheden om bestanden te combineren.

#### 3.3.1 Koppelen en het doel van de koppeling van bronnen

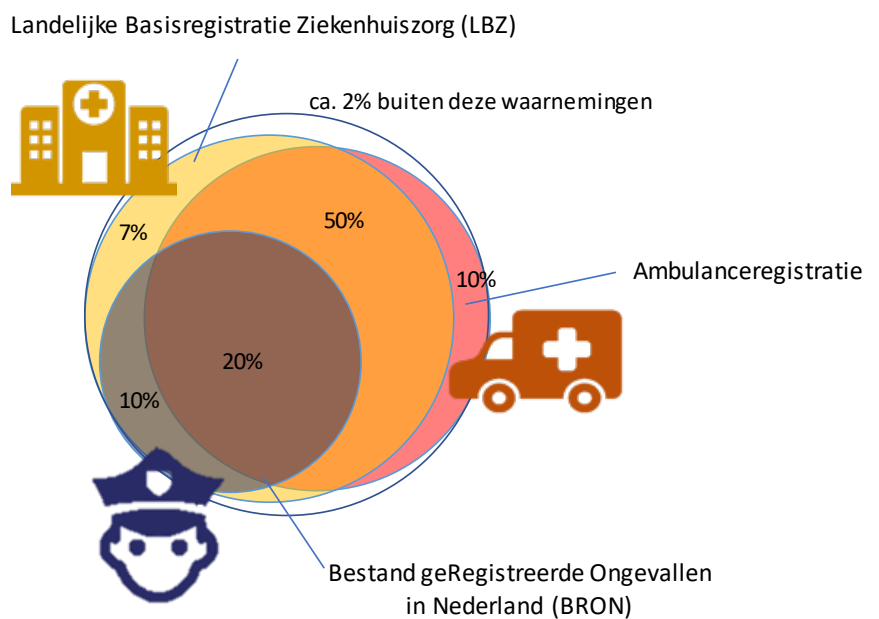
Er zijn grofweg drie mogelijke doelen voor een koppeling van databestanden:

- Verrijking van een bestand (bijvoorbeeld: BRON aanvullen met behulp van ambulancegegevens; er komen dan bijvoorbeeld méér ongevallenlocaties en slachtoffers in beeld; er zullen echter ook nog ongevallen of slachtoffers ontbreken).
- Aanvullingen op de vaststelling van de werkelijke omvang van een probleem, zoals het aantal ernstig verkeersgewonden. In dit geval gaat het niet alleen om koppelen, maar ook om de vaststelling van het totale aantal slachtoffers door 'bijschatting' toe te passen op dat wat logischerwijze ontbreekt in de gekoppelde bestanden (zie *Afbeelding 1* en *Paragraaf 3.3.3*).
- Een combinatie van bovenstaande, waarbij vooral meer kenmerken van werkelijke aantallen slachtoffers, maar ook ongevallen kunnen worden onderzocht en gepubliceerd. Door ambulancegegevens toe te voegen aan de koppeling van BRON en LBZ, kan bijvoorbeeld van een groter deel van de ongevallen en slachtoffers de ongevalslocatie worden bepaald. Ook zijn naar verwachting meer disaggregaties van kenmerken van werkelijke aantallen mogelijk.

a.  
**Overlap in letselongevallen tussen BRON en ambulanceregistratie**



b.  
**Verwachte overlap in registraties met betrekking tot ernstig verkeersgewonden**



**Afbeelding 1a en 1b.** Venndiagrammen die laten zien hoe de verschillende registraties elkaar overlappen en aanvullen: de bovenste afbeelding (a) geeft schematisch de overlap weer tussen de letselongevallen in BRON en de verwachte overlap met de ambulanceregistratie; de onderste afbeelding (b) geeft aan hoe naar verwachting BRON, de LBZ en de ambulanceregistratie zullen overlappen (zie Bijlage G voor de achterliggende onderbouwing).

Voor beleidsmakers is vooral het eerste doel (meer slachtoffers en ongevalslocaties in beeld) aan te sluiten bij wat zij voor ogen hebben (zie ook *Hoofdstuk 2* van deze notitie): afzonderlijke bestanden zijn interessant, maar door bestanden met elkaar te combineren en bij voorkeur te koppelen, kan duidelijker worden hoe registraties zich tot elkaar verhouden. Dit betekent dat het ambulancebestand geschikt moet zijn om in ieder geval met BRON te kunnen koppelen en zo gedetailleerd (gedisaggregeerd) mogelijk beschikbaar te stellen. Op die manier weten wegbeheerders waar ongevallen die momenteel (nog) niet in BRON zitten hebben plaatsgevonden en wat de kenmerken van deze ongevallen waren. Ook een koppeling met de LBZ kan duidelijk maken wat de locatie is geweest van ongevallen waardoor betrokkenen in het ziekenhuis zijn opgenomen.

Het tweede en derde doel is vooral op een wat hoger niveau relevant: het kan met name op nationaal niveau maar ook op regionaal niveau een beter beeld verschaffen van het totale aantal ernstig verkeersgewonden. Het gaat dan vooral om de kenmerken van de slachtoffers en ongevallen waarin deze slachtoffers letsel hebben opgelopen. Hierbij komt vooral meer kennis beschikbaar over de locatie waar deze ongevallen hebben plaatsvonden, maar ook hoe slachtoffers en ongevallen aan elkaar gerelateerd zijn, in aanvulling op wat daarover uit BRON blijkt. Beleidsmakers kunnen deze kennis benutten om beleid te formuleren dat aansluit bij verkeersveiligheidspatronen in Nederland. Zo kan uit een landelijke analyse van ernstig verkeersgewonden met aanvullende informatie van ambulancegegevens bijvoorbeeld blijken dat bepaalde typen slachtoffers vooral op bepaalde locaties een ongeval hebben gehad. Op basis van die kennis kunnen wegbeheerders bekijken of dit binnen hun beheergrenzen relevant kan zijn en welke problemen daaraan ten grondslag liggen. Op basis daarvan kunnen ze vervolgens adequate beleidsoplossingen formuleren en implementeren. Dat kan bijvoorbeeld ook bijdragen aan betere regionale schattingen en betere schattingen ten aanzien van specifieke groepen slachtoffers (bijvoorbeeld voor bromfietzers of oudere fietsers).

### 3.3.2 Beschikbare data en wat dit betekent voor koppelmogelijkheden

Een algemeen uitgangspunt bij het combineren van bestanden is dat de kwaliteit van de combinatie toeneemt naarmate er meer zekerheid is of een regel met gegevens in het ene bestand op hetzelfde voorval of persoon betrekking heeft als een regel met soortgelijke gegevens in het andere bestand. Dat wil zeggen dat in een kwalitatief hoogwaardige combinatie van bestanden er een hoge mate van zekerheid is dat een slachtoffer in bijvoorbeeld BRON hetzelfde slachtoffer is dat ook in de ambulanceregistratie gevonden wordt met vergelijkbare kenmerken. Is die zekerheid hoog, dan spreken we van koppelen.

Hiervoor is het van belang dat de informatie die wordt benut om bestanden te combineren, voldoende *onderscheidend vermogen* heeft. Dat wil zeggen dat als ongevals- of slachtofferkenmerken slechts een beperkt aantal mogelijke waarden hebben (zoals het kenmerk 'geslacht' of het kenmerk 'wegvak/kruispunt'), het dan veel lastiger is om er zeker van te zijn dat een bepaald slachtoffer of ongeval in het ene bestand op basis van dit kenmerk hetzelfde is als in het andere bestand. Er zijn in het slechtste geval dan alleen uitspraken over totalen of grove aantallen ongevallen met bepaalde kenmerken mogelijk (zie ook *Bijlage B*). Een grotere koppelkwaliteit kan bereikt worden door één of meer koppelvariabelen ter beschikking te hebben die personen of voorvallen zo uniek mogelijk kunnen identificeren. Dit kan bijvoorbeeld door een uniek identificatienummer te benutten zoals het burgerservicenummer (BSN), of een combinatie van kenmerken die voldoende onderscheidend vermogen hebben. Wat hierbij 'voldoende' is, hangt mede af van de vraag en het vóórkomen van de gebruikte koppelkenmerken in de te koppelen populatie.

De koppeling tussen bronnen, het proces dat dit vergt en de organisaties die hierbij betrokken zijn, hangen ook samen met de beschikbare data van de verschillende bronnen, de mate waarin de data tot personen te herleiden zijn en de mate waarin ambulanceorganisaties de herleidbaarheid als zodanig beoordelen.

Grofweg zijn de volgende scenario's te onderscheiden:

- > koppeling op basis van BSN;
- > koppeling op basis van andere persoons- en ongevalskenmerken;
- > koppeling zonder persoonskenmerken.

### **BSN-koppeling**

Bestanden die op basis van het BSN gekoppeld kunnen worden, zijn:

- de ambulancegegevens;
- de ruwe politieregistratie;
- de LBZ.

Als het gaat om koppelen op basis van persoonsinformatie, biedt een BSN-koppeling in principe het beste resultaat. Zoals we in *Paragraaf 3.2.2* zagen, gelden er voor de verwerking van burgerservicenummers wel strenge restricties en moet de informatie versleuteld worden via een Trusted Third Party (TTP).

Om de ruwe politiegegevens te mogen gebruiken, moet het ministerie van Justitie en Veiligheid (JenV) per aanvraag toestemming geven aan de politie om hieraan mee te werken, en aan verwerkers van de data om deze te mogen gebruiken. Een BSN- koppeling met de politieregistratie kan op dit moment alleen worden uitgevoerd met de ruwe politiedata (en niet met BRON, omdat het BSN daarin ontbreekt). Daarom lijkt deze variant onder de huidige condities vooral interessant voor jaarlijkse schatting van het aantal ernstig verkeersgewonden. Daarnaast heeft deze koppelvariant ook potentie voor onderzoek naar de verbetermogelijkheden van BRON. Op termijn zou een manier gevonden kunnen worden waarop ook (versleutelde) uitlevering van BSN-gegevens aan Rijkswaterstaat ter verrijking van BRON mogelijk is. Dit zou met name de voorkeur hebben van beleidsmakers die graag een centraal ongevalbestand aanhouden. BRON is hiervoor de basis en zal dat naar verwachting ook blijven.

### **Gecombineerde persoons- en ongevalsgegevenskoppeling**

Koppeling van databronnen op basis van kenmerken anders dan BSN-gegevens, wordt al elk jaar toegepast bij de vaststelling van het aantal ernstig verkeersgewonden, waarbij SWOV een koppeling maakt tussen BRON en de LBZ. Hiervoor is verdere versleuteling niet nodig, mits de verwerking van de gegevens door onderzoekers in een beveiligde omgeving plaatsvindt waarvoor de betrokken data-eigenaren hun toestemming hebben gegeven (zoals in dit geval bij het CBS). De resultaten van deze verwerking die buiten deze omgeving worden gebracht, mogen nooit terug te herleiden zijn naar personen (en in het geval van het CBS ook niet naar instellingen).

In hoeverre ambulancegegevens op eenzelfde wijze verwerkt mogen worden, met dezelfde mate van detail over betrokken personen als de LBZ, moet nog nader worden bezien met de regionale ambulancevoorzieningen (RAV's). Mogelijk kan de wijze waarop SWOV jaarlijks de LBZ verwerkt om het aantal ernstig verkeersgewonden vast te stellen, gebruikt worden als inspiratie voor de RAV's om ook ambulancegegevens veilig beschikbaar te stellen. De verwachting is dat de visies over wat wel of niet acceptabel is en de condities waaronder, verschillen per RAV<sup>16</sup>. Het ligt in de rede om hier op basis van een gezamenlijk met de centrale overheid uitgewerkt voorstel verder het gesprek over aan te gaan met de ambulancevoorzieningen en hun koepelorganisaties AZN en GGD GHOR.

In *Bijlage F* is aan de hand van de casus Flevoland meer inzicht gegeven in de mogelijkheden bij een van de ambulancevoorzieningen. Deze ambulancevoorziening is tot nu toe de enige die



16 Zo heeft SWOV onder voorwaarden van de GGD Flevoland data ontvangen voor onderzoek dat meer gedisaggregeerd was dan de open data. In eerder onderzoek dat SWOV samen met RIVM uitvoerde in 2015 verleenden alle RAV's hun medewerking met verschaffen van data, behalve de RAV Groningen die hier problemen mee had vanwege mogelijke aantasting van de vertrouwelijkheid van gegevens.

ambulancedata over verkeersongevallen als open data beschikbaar heeft gesteld, en daarnaast meewerkt aan onderzoek waarbij ambulancedata zijn gecombineerd met data uit BRON en de LBZ om zo een beter beeld te krijgen van het aantal ernstig verkeersgewonden in een regio (Wijlhuizen & Bos, te verschijnen). Hieruit blijkt dat uitlevering van meer details over slachtoffers aan onderzoeksinstanties mogelijk is. Dit beeld wordt ook bevestigd door een recente studie van VeiligheidNL met ambulancedata van de Regionale Ambulance Voorziening Utrecht (RAVU; Olij & Nijman, 2020). Het beeld dat hieruit naar voren komt, is dat een aantal RAV's aan onderzoeksinstanties onder voorwaarden ook iets specifiekere data willen uitleveren over met name betrokken ongevalsslachtoffers. Het gaat hierbij overigens nog steeds om geaggregeerde (geclusterde) data (zoals de leeftijd van slachtoffers in klassen). Bij deze koppelingen kunnen naast overeenkomsten in slachtofferinformatie ook overeenkomsten in locatie-informatie worden gebruikt (net als bij de open data, zie hierna). Daarnaast is het voor het combineren met de LBZ van waarde om te weten naar welk ziekenhuis het slachtoffer is vervoerd en wat de aard van de inzet was (behandeld ter plaatse of vervoerd naar ziekenhuis). Overigens moet opgemerkt worden dat zonder nog verder gespecificeerde informatie over slachtoffers (met name leeftijd in jaren met zo mogelijk een toevoeging van maand of dag) een echte koppeling met regels uit andere bestanden niet mogelijk lijkt.

### **Koppeling zonder persoonsgegevens**

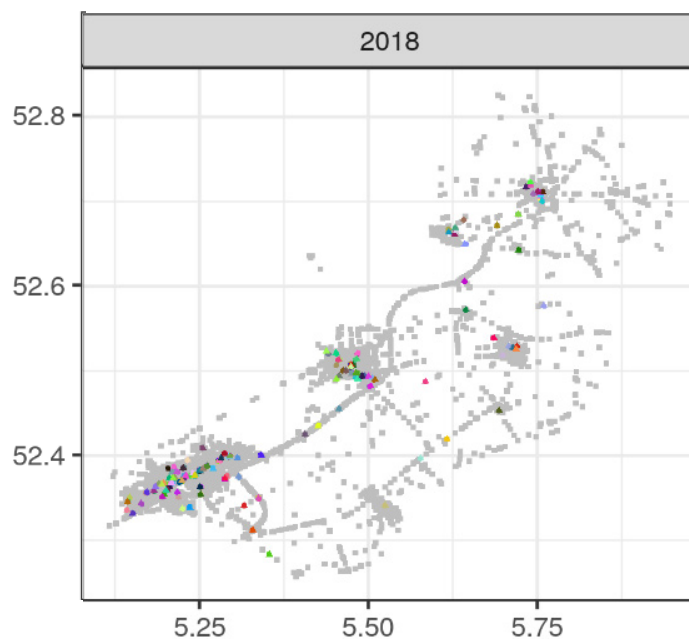
De minst gevoelige variant om databronnen te combineren, betreft die op basis van data zonder tot personen herleidbare gegevens. Dit zijn data waar in principe vanuit de privacywetgeving geen restricties voor gelden en die openbaar voor iedereen beschikbaar kunnen worden gesteld (open data). Op dit moment is dat voor ambulancegegevens alleen nog maar mogelijk met de data van de GGD Flevoland, die een aggregaat van haar ambulanceritten naar aanleiding van verkeersongevallen als open data beschikbaar heeft gesteld<sup>17</sup>.

Een combinatie tussen ongevallenregistraties in verschillende bestanden vindt in deze variant plaats op basis van overeenstemming in ongevalsjaar, locatie van het ongeval en eventueel de vervoerswijze van slachtoffer en tegenpartij. Omdat de beschikbare data weinig onderscheidend zijn (zoals specifiekere ongevalsdatum, leeftijd en geslacht van het slachtoffer), kan verwacht worden dat de koppelkwaliteit niet hoog zal zijn (we laten hieronder een voorbeeld zien). Voor combinatie met de LBZ kan de locatie niet benut worden en is alleen een vergelijking op aantallen mogelijk. Om een daadwerkelijke koppeling tot stand te brengen, zijn dus meer gegevens noodzakelijk dan de nu beschikbare open data.

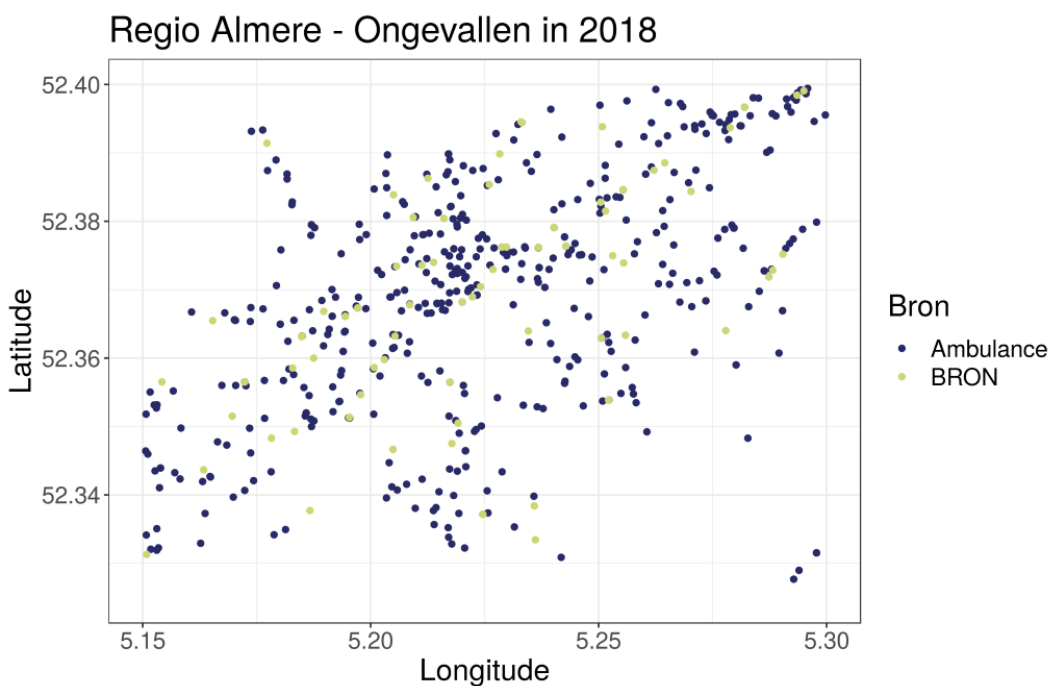
De *Afbeeldingen 2, 3 en 4* tonen respectievelijk de combinatiemogelijkheden van BRON (ongevallen met ten minste één letselslachtoffer) met de open ambulancedata van Flevoland geplot op een kaart (*Afbeelding 2*) en ingezoomd op een gemeente (*Afbeelding 3*) als we deze gegevens op een deterministische wijze proberen te koppelen (zie *Paragraaf 3.3.3* en *Bijlage H* voor uitleg over verschillende koppeltechnieken en resultaten als we deze toepassen op de genoemde open data). *Afbeelding 2* laat zien dat ongeveer een kwart van de ambulancegegevens gekoppeld kan worden aan een BRON-gegeven. Dit aandeel kan verhoogd of verlaagd worden door respectievelijk grotere of kleinere afstanden tussen de locaties van BRON-gegevens en ambulancegegevens toe te staan, maar dit heeft ook gevolgen voor het aantal (in)correcte combinaties dat gemaakt zal gaan worden. We hebben immers geen zekerheid over welke BRON-ongevallen bij welke ambulanceritten horen, maar proberen dat door koppeling van overeenkomstige gegevens zo goed mogelijk te weten te komen.



<sup>17</sup> <https://www.eengezonderflevoland.nl/verkeersongevallen-flevoland-almere-lelystad> of <https://data.overheid.nl/dataset/verkeersongevallen-met-ambulancezorg-in-flevoland>



**Afbeelding 2.** Gevonden combinaties tussen het ongevalbestand BRON (letselongevallen) en de open ambulancedata van de GGD Flevoland van verkeersongevallen in 2018. De gekleurde punten geven combinaties weer, de grijze punten niet-gecombineerde gegevens.

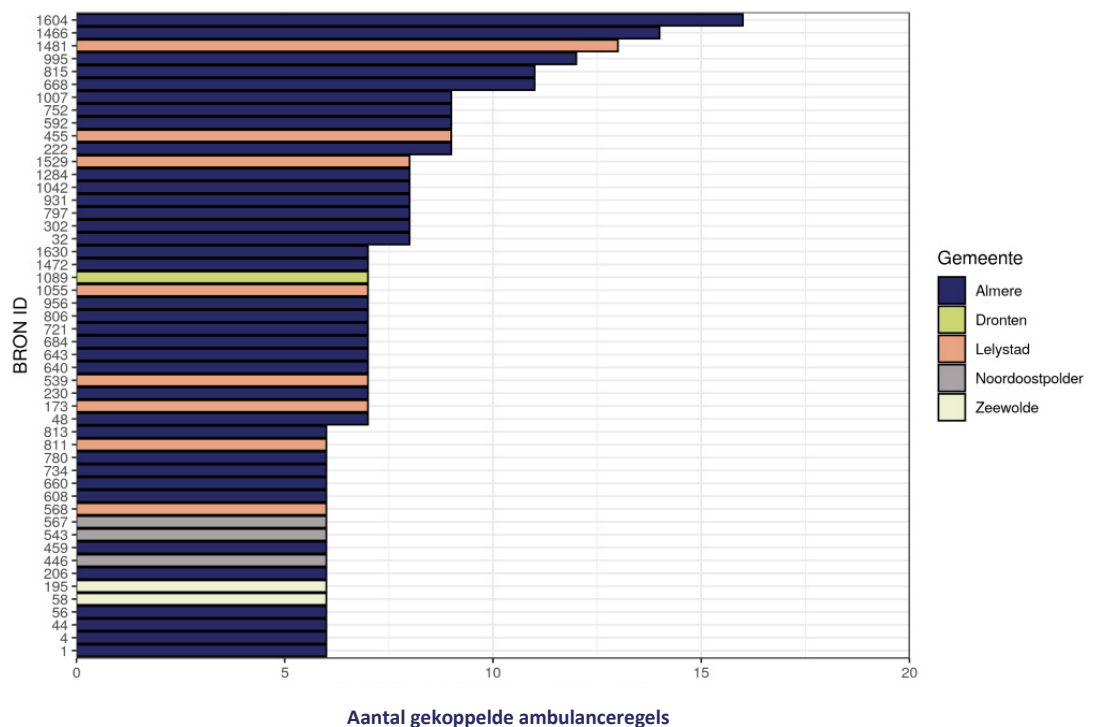
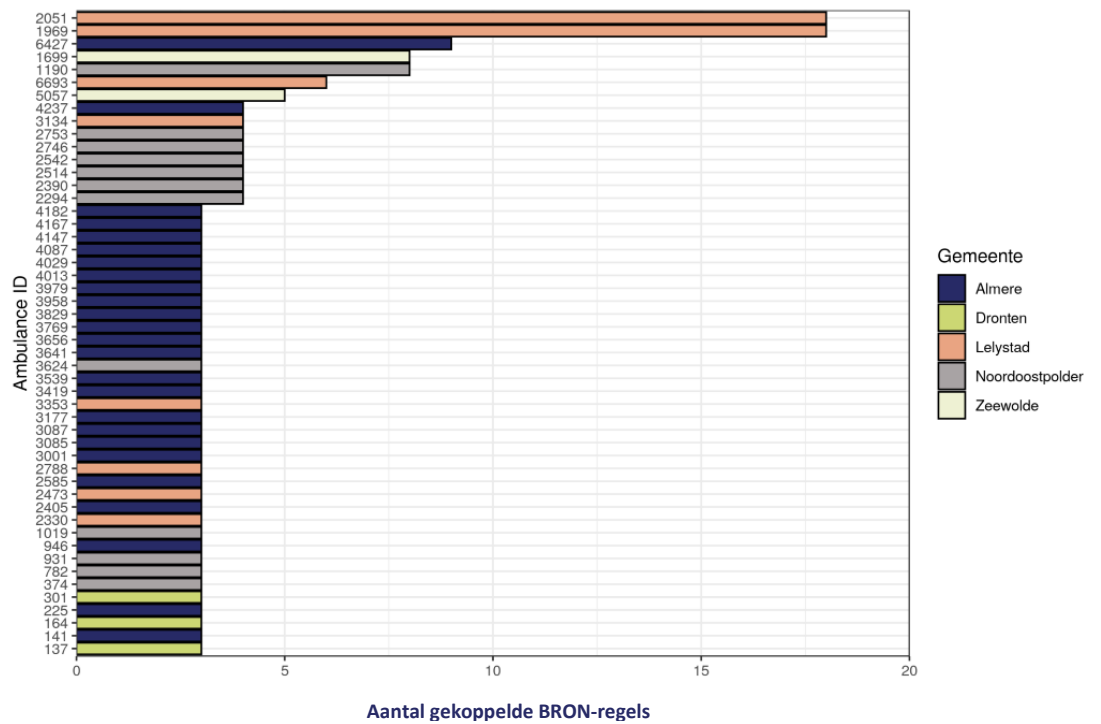


**Afbeelding 3.** Locaties van letselongevallen volgens BRON en locaties volgens ambulanceritten in de open data van GGD Flevoland voor de gemeente Almere in 2018.

De analyse laat zien dat de beschikbare gegevens beperkt onderscheidend zijn, waardoor het niet altijd mogelijk is om elk gegeven uit de ene bron uniek te koppelen met een gegeven uit de andere bron. Dit wordt nog verder duidelijk in *Afbeelding 4*, waarin aangegeven staat hoe vaak een regel uit de ene bron gekoppeld wordt met een regel uit de andere bron. De afbeelding toont de vijftig meest voorkomende combinaties, geordend naar het aantal gemaakte koppelingen. Zo kan een bepaalde BRON-registratie (bijvoorbeeld regel 1604) met meer dan vijftien ambulanceregels worden gekoppeld die overeenkomen in locatie (en dus ook gemeente, zoals in de afbeeldingen te zien is) en ook in hetzelfde jaar geregistreerd zijn.



Anders gezegd: door gebrek aan onderscheidende kenmerken worden hier slechts clusters van ongevallen gevonden die in kenmerken overeenkomen. Hierdoor is niet duidelijk welk deel van deze overeenkomstige ongevallen daadwerkelijk overlapt of aanvullend is op de ongevallen in het andere bestand.



**Afbeelding 4.** De vijftig meest voorkomende combinaties indien de ambulancedata als uitgangspunt worden genomen (bovenste afbeelding) en als BRON (alleen letselongevallen) als uitgangspunt wordt genomen (onderste afbeelding) (zie Bijlage H voor uitleg over de wijze waarop deze afbeeldingen tot stand zijn gekomen).

### 3.3.3 Technieken om bestanden te koppelen en werkelijke omvang vast te stellen

Om een beter beeld te krijgen van het werkelijke aantal ongevallen en slachtoffers en de bijbehorende kenmerken, combineren we in Nederland verschillende bestanden (voor de verkeersdoden wordt dat gedaan door het CBS, voor ernstig verkeersgewonden door SWOV). Daarnaast kan het nodig zijn nog extra correcties uit te voeren om een goed beeld van de werkelijkheid te krijgen. Hieronder gaan we kort in op deze technieken die van belang zijn als ambulancedata worden toegevoegd aan andere bestanden zoals BRON of de LBZ.

Er zijn verschillende manieren om gegevensbestanden te koppelen. Welke dat zijn, is mede afhankelijk van de mate waarin de te koppelen bestanden unieke identificatiekenmerken bevatten (zoals een BSN) en ook van de kans op fouten of onnauwkeurigheden bij de registratie of andere verklaringen (bijvoorbeeld een tijdsverschil tussen de datum en tijdstip van het ongeval en de ziekenhuisopname) waardoor gegevens van eenzelfde ongeval of persoon niet helemaal overeen kunnen komen maar wel op hetzelfde betrekking hebben.

Hieronder gaan we in op de verschillende koppeltechnieken en hoe deze aansluiten bij verschillende situaties:

- deterministisch koppelen;
- koppelen door middel van een afstandsfunctie;
- probabilistisch koppelen.

Als laatste gaan we ook in op een procedure die aanvullend is op koppeling van bestanden om door middel van bijschatting de werkelijke omvang van het aantal verkeersslachtoffers vast te kunnen stellen. Deze stap is niet nodig als het doel is om bijvoorbeeld BRON te verrijken met ambulancegegevens. Daarvoor volstaat toepassing van met name de eerste of tweede koppeltechniek. Koppeling en vervolgens bijschatting wordt wel toegepast bij bijvoorbeeld de vaststelling van het aantal ernstig verkeersgewonden (het gaat daarbij dan om vaststelling van een werkelijke omvang, wat meer is dan de combinatie van twee bestanden) en is dus vooral in dat proces relevant, ook als daar ambulancegegevens in worden betrokken.

#### Deterministisch koppelen

Een relatief eenvoudige techniek om regels in bestanden aan elkaar te koppelen is door te bepalen welke regels met gegevens daadwerkelijk overeenkomen met alle of een gedeelte van de te koppelen kenmerken (bijvoorbeeld de geboortedatum en datum van het ongeval, maar ook in geval van een gepseudonimiseerd BSN). Hiermee zijn nog niet alle problemen opgelost. Zo zullen ontbrekende gegevens van slachtoffers in de verschillende registraties ervoor zorgen dat regels niet aan elkaar gekoppeld kunnen worden. Wordt bijvoorbeeld de geboortedatum van het slachtoffer gebruikt om te koppelen maar ontbreekt deze in een van de bestanden, dan is het onzeker of het om dezelfde persoon gaat. Ook fouten in gegevens (bijvoorbeeld door invoerfouten) kunnen ervoor zorgen dat de gegevens bij deterministisch koppelen niet gekoppeld worden. Door de gegevens te onderzoeken op onwaarschijnlijke gegevens (zoals een leeftijd van ver boven de 100 jaar), kan worden bepaald welke waarden waarschijnlijk incorrect zijn. Daarmee is echter de juiste waarde (in dit geval de leeftijd of geboortedatum) nog niet bekend.

Bij ongevallenregistraties worden veelal ook de datum en tijdstip van het ongeval gebruikt om koppelingen met andere bestanden te bepalen. Het tijdstip van het ongeval (bijvoorbeeld in BRON of in het ambulancebestand) zal echter vrijwel nooit dezelfde zijn als die van een ziekenhuisopname (in de LBZ). Door in het koppelproces te vereisen dat deze gegevens exact overeenkomen, zullen de regels die in de verschillende bronnen eigenlijk bij elkaar horen, onterecht niet gekoppeld gaan worden. Dit speelt vooral bij een bepaald soort letsel, of wanneer er grotere afstand zit tussen ongevals- en opnamelocatie (zie bijvoorbeeld Bos et al., 2019). Daarom passen we deze methode binnen de verkeersveiligheid liever niet toe maar maken we gebruik van een iets geavanceerdere techniek, die hierna wordt beschreven.

- Voordelen: eenvoudig te begrijpen en toe te passen. Alleen overeenkomende gegevens worden aan elkaar gerelateerd.
- Nadelen: laat geen registratiefouten of -verschillen toe die in werkelijkheid wel over hetzelfde gaan.

### **Koppelen op basis van een afstandsfunctie**

Een alternatief voor deterministisch koppelen is het gebruik van een 'afstandsfunctie', die een gewogen gemiddelde berekent tussen de virtuele 'afstanden' die tussen de verschillende gegevens kan worden vastgesteld. Zo kan de afstand tussen twee tijdstippen worden bepaald door te kijken hoeveel eenheden er zitten om de dag, maand en jaar te laten overeenkomen, waarbij een perfecte overeenkomst een afstand = 0 oplevert en bijvoorbeeld twee data met één dag verschil een afstand = 1 opleveren. Daarbij kan ook nog worden bepaald of de afstand positief (het opnametijdstip van het ziekenhuis ligt altijd na het ongeval) of negatief (het opnametijdstip ligt vóór het ongeval) is.

Vervolgens kunnen deze afstanden worden gewogen (bijvoorbeeld onder de aanname dat fouten bij het schrijven van een tekst waarschijnlijker zijn dan bij het schrijven van een datum en juist minder aannemelijk als we zouden weten dat bepaalde informatie wordt ingescand in plaats van met de hand opgeschreven) en worden gecombineerd tot een getal: de gewogen afstand. Gegevens uit twee bronnen worden dan gecombineerd als deze gewogen afstand minder dan een vooraf vastgestelde waarde is. Wat een redelijke waarde is om nog een terechte koppeling te mogen veronderstellen, vergt afstemming met wat er in de praktijk zoal kan voorvallen (bijvoorbeeld: hoe waarschijnlijk is het dat iemand die in Maastricht een ongeval krijgt in Groningen wordt opgenomen, zeker als er ook redelijk wat tijd tussen ongeval en opname zit?).

Deze koppelmethode op basis van een afstandsfunctie wordt onder andere toegepast bij de vaststelling van het aantal ernstig verkeersgewonden in Nederland.

- Voordelen: maakt het mogelijk geavanceerdere en conditie-afhankelijke relaties tussen gegevens te beschrijven en foutenmarges toe te laten.
- Nadelen: houdt geen rekening met de kans van vóórkomen van bepaalde combinaties van kenmerken op basis waarvan de waarschijnlijkheid bepaald kan worden dat twee registraties inderdaad op hetzelfde slaan.

### **Probabilistisch koppelen**

De methode met de afstandsfunctie heeft een belangrijke beperking: ze houdt geen of weinig rekening met kansen dat bepaalde fouten optreden, of met kansen dat twee slachtoffers op dezelfde dag een ongeval hebben en op dezelfde dag zijn geboren. Ook is er geen eenduidige manier om de gewichten voor de weging van de verschillende afstanden vast te stellen. Om het belang van de verschillende variabelen op basis van de gegevens te schatten en rekening te houden met de verschillende relevante kansen, kan een derde methode worden gebruikt, die bekend staat als 'probabilistisch koppelen' (Christen, 2012). In deze methode wordt als eerste ingeschat hoe groot de kans is dat bijvoorbeeld de geboortedatum overeenkomt als gegevens uit twee bronnen wel over dezelfde persoon gaan (de zogenaamde 'm-kans') en hoe groot de kans is dat dezelfde geboortedatum overeenkomt als de gegevens niet over dezelfde persoon gaan (de zogenaamde 'u-kans'). Het schatten van deze kansen kan op verschillende manieren worden gedaan. Indien de kansen worden geschat op basis van beschikbare gegevens zonder dat zeker is dat deze gegevens bij elkaar horen, is het van belang dat de gegevens niet te rommelig zijn en niet te veel fouten bevatten. Probabilistisch koppelen heeft vooral als voordeel dat voor elke paar gegevens een kans beschikbaar komt die aangeeft hoe waarschijnlijk het is dat deze gegevens van dezelfde persoon afkomstig zijn. De eerder genoemde afstandsfunctie kan een vergelijkbare functie vervullen, maar daarbij is het mogelijk lastiger in te schatten wat een grote of kleine waarde precies betekent. Probabilistisch koppelen wordt momenteel nog niet binnen de

verkeersveiligheid toegepast, maar SWOV verkent of het interessant is om dat in de toekomst wel te gaan doen.

- Voordelen: maakt het mogelijk op basis van mate van vóórkomen van kenmerken de kans te bepalen dat bepaalde registraties bij elkaar horen. Daarbij is geen daadwerkelijke koppeling meer nodig maar kan bijvoorbeeld het vóórkomen van combinaties van kenmerken in ongevallen beter worden geschat.
- Nadelen: complexe methode die nog niet binnen de verkeersveiligheid wordt toegepast. Is alleen aantrekkelijk als koppeling op andere wijzen niet goed mogelijk is of onvoldoende mogelijk maakt om onderliggende combinaties van kenmerken te beschrijven.

### **Bijschatten van aantallen: nodig om werkelijke aantallen te bepalen**

Niet alle verkeersongevallen worden geregistreerd of zijn direct als verkeersongeval in een registratie herkenbaar (zie eerder). Er is dus sprake van onderregistratie. Aangezien geen enkele bron op zichzelf een compleet beeld geeft van alle verkeersongevallen, is bij alle bronnen dus sprake van onderregistratie of ten minste gedeeltelijke misclassificatie, waardoor verkeersongevallen of verkeersslachtoffers niet als zodanig herkenbaar zijn. Door bronnen te koppelen kan een completer beeld worden verkregen van de werkelijke omvang van het aantal ongevallen of slachtoffers: de bronnen vullen elkaar in dat geval op onderdelen aan. Dat wil echter nog niet zeggen dat daarmee het beeld helemaal compleet is. Het kan immers nog steeds zo kan zijn dat er ongevallen of slachtoffers in geen van de gekoppelde bronnen voorkomen. Dit probleem is reeds aangepakt voor de koppeling tussen BRON en de LBZ om jaarlijks het aantal ernstig verkeersgewonden in Nederland te bepalen (Reurings & Bos, 2009; Bos et al., 2019). Hierbij wordt als uitgangspunt genomen dat het ene bestand niet compleet is (BRON) en dat in het andere bestand niet alle *verkeersslachtoffers* herkenbaar zijn (de LBZ). Door te bepalen a) welk aantal slachtoffers in beide bestanden herkenbaar zijn als verkeersslachtoffer en b) welk aantal slachtoffers wel in de ene maar niet in de andere bron aanwezig of herkenbaar is, kan bepaald worden welk aantal slachtoffers in beide bronnen ontbreekt of niet herkenbaar is als verkeersslachtoffers. Deze laatste stap wordt de ‘bijschatting’ genoemd en kan een relevante stap zijn bij de combinatie van bronnen als hierbij het doel is om een schatting van de werkelijke omvang van het aantal verkeersslachtoffers te bepalen.

### **Conclusies ten aanzien van koppeltechnieken**

Voor verrijking van BRON zijn deterministische koppelmethoden, bij voorkeur met afstandsfunctie aan te bevelen. Hierbij wordt naar zo veel mogelijk overeenkomst in kenmerken van slachtoffers en/of ongevallen gekeken om een koppeling tussen regels van verschillende bestanden vast te stellen.

Voor verrijking van ernstig verkeersgewonden wordt vooralsnog vastgehouden aan koppeling met afstandsfunctie. Op termijn is eventueel een probabilistische koppeltechniek interessant, waarbij gekoppeld wordt op basis van kansen van het vóórkomen van kenmerken, maar dit wordt nog nader verkend. Een extra stap bij de vaststelling van werkelijke aantallen is de bijschattingsprocedure, waarbij wordt vastgesteld welk deel van slachtoffers of ongevallen in de gekoppelde bestanden nog ontbreekt. Deze extra stap naast de koppeling is al een vast onderdeel bij de vaststelling van het aantal ernstig verkeersgewonden in Nederland.

## 3.4 Samenvattend: koppelmogelijkheden en vereisten vanuit de data

Er blijken verschillende mogelijkheden om databronnen aan elkaar te koppelen (of beter: met elkaar te combineren), afhankelijk van de wensen van eindgebruikers (zie *Hoofdstuk 2*), de gebruikte databronnen, de daarin voorkomende variabelen en detailniveau daarvan, maar ook de interpretatie van gegevensbeschermingswetgeving. Al met al kan dit als volgt worden samengevat, waarbij we per koppelbron stilstaan bij de mogelijkheden en consequenties, alsmede bij de betrokken partijen en hun rol.

### 3.4.1 Verrijking van BRON

Hier blijken in theorie twee mogelijkheden beschikbaar:

1. combinatie met open data;
2. combinatie via specifieke uitlevering van bronnen.

#### Combinatie met open data

Dit betreft de mogelijkheid om ongevallen uit BRON (slachtoffers zijn momenteel niet als open data beschikbaar) te combineren met de open data van ambulanceritten. Vooralsnog is dit alleen mogelijk voor data die betrekking hebben op de provincie Flevoland, omdat dit momenteel de enige ambulancevoorziening is die een voor iedereen toegankelijk databestand beschikbaar heeft gesteld (open data). Alhoewel een combinatie van data op algemene kenmerken en vooral locaties van ongevallen mogelijk is, leidt dit slechts tot een grove combinatie van bestanden waarbij geen een-op-een-onderscheid tussen gecombineerde ongevallen mogelijk is. Dit komt doordat gegevens zodanig zijn geanonimiseerd dat ongevallen en slachtoffers onvoldoende onderscheidbaar zijn. Dit geldt vooral voor betrokken slachtoffers. Deze mogelijkheid lijkt voor de verrijking van BRON in praktische zin daarom niet echt interessant.

#### Betrokken partijen

In dit geval zijn er geen specifieke betrokken partijen anders dan de bronleveranciers die besluiten of en welke data als open data beschikbaar worden gesteld. Koppelingen zijn door iedereen te maken.

#### Combinatie van specifieke uitlevering van bronnen

Uit een aantal casussen (Flevoland en Utrecht) blijkt het recentelijk mogelijk om ook in iets meer detail gegevens van ambulanceritten aan onderzoeksinstanties uit te leveren en in de data een tussenweg te vinden in wat als ‘anonieme gegevens’ beschouwd kunnen worden en wat zich dan toch beter kan lenen voor combinaties met bijvoorbeeld de LBZ en BRON (ook hierbij is meer detailniveau dan de open data vereisen). Hierbij is bijvoorbeeld de leeftijd van slachtoffers in klassen met toevoeging van geslacht mogelijk, evenals informatie over het ziekenhuis waarnaar het slachtoffer vervoerd is en de aard van de rit (behandeld ter plaatse of vervoerd naar ziekenhuis). Dit laatste is vooral van waarde voor een combinatie met de LBZ.

Hierbij dient te worden opgemerkt dat bij geen van deze casussen met de beschikbare data een daadwerkelijke koppeling van regels in verschillende bestanden mogelijk was. Hiervoor zijn nog gedetailleerdere data nodig. Indien de leeftijd van het slachtoffer inclusief toevoeging van maand of dag nog te veel herleidbaarheid geeft, lijkt de enige weg om persoonsgegevens van slachtoffers in bestanden te versleutelen zodat de gegevens niet meer herleidbaar zijn, maar wel uniek onderscheidbaar. Deze mogelijkheid lijkt op korte termijn het meest kansrijk om verder te verkennen. Koppelingen kunnen het beste tot stand worden gebracht door toepassing van deterministische koppeltechnieken, liefst met afstandsfunctie, om zo voor kleine verschillen in tijdstippen en eventuele registratiefoutjes te verdisconteren.

#### *Betrokken partijen*

Bij onderzoeksgegevens die worden geanonimiseerd door de bronhouder zelf, zijn de bronhouders (RAV's, Rijkswaterstaat voor BRON en DHD voor de LBZ) en onderzoekers de betrokken partijen. Zij maken afspraken over wat de noodzaak is van uitlevering van gegevens, wat het nodige en mogelijke detailniveau is en komen overeen wat en hoe er onder welke voorwaarden wordt uitgewisseld.

Indien het wenselijk is om een daadwerkelijke koppeling van gegevens te maken, dan is tussenkomst van een vertrouwde derde partij (TTP) nodig die ervoor zorgt dat de personen in de te koppelen bestanden op unieke wijze worden versleuteld zodat ze niet meer te herleiden zijn maar nog wel uniek onderscheidbaar van elkaar zijn.

### **3.4.2 Koppeling met ruwe politiegegevens**

Een variant die we in deze notitie apart hebben besproken, is de mogelijkheid om ambulancegegevens te koppelen met ruwe politiegegevens. Dit heeft als belangrijke voordelen ten opzichte van BRON dat er BSN-gegevens van slachtoffers beschikbaar zijn, maar ook dat er meer geregistreerde ongevallen beschikbaar zijn die om wat voor reden dan ook niet direct als verkeersongeval zijn geclassificeerd maar dat wel zijn. Dit traject betekent wel dat er meer partijen bij betrokken moeten worden. Het lijkt vooralsnog vooral een traject dat interessant is om te benutten voor verder onderzoek en dat op termijn tot interessante aanvullingen van BRON kan leiden.

#### *Betrokken partijen*

Ten eerste zal bij het ministerie van Justitie en Veiligheid toestemming moeten worden verkregen om ruwe politiegegevens voor koppeldoelinden beschikbaar te maken. Zeker als dit op structurele basis gaat gebeuren, moet goed nagedacht worden over een wijze waarop de politie niet extra wordt belast. Vervolgens moeten met de bronhouders (politie en ambulancevoorzieningen) afspraken gemaakt worden over wat precies uitgeleverd mag worden en onder welke voorwaarden. Omdat de meerwaarde van deze variant vooral zit in benutting van het BSN, is in dit geval zeker de betrokkenheid van een vertrouwde derde partij (TTP zoals zorgTTP of CBS) noodzakelijk die ervoor zorgt dat de gegevens versleuteld worden. Uiteindelijk dienen met de ontvangende partij (vooralsnog is dat alleen voorbehouden aan onderzoekers) afspraken gemaakt te worden over de precieze voorwaarden van verwerking.

### **3.4.3 Koppeling met de LBZ**

In geval van koppeling met de LBZ, bijvoorbeeld om informatie over ernstig verkeersgewonden te verkrijgen, zijn er twee scenario's:

1. koppeling op basis van BSN-gegevens;
2. koppeling op basis van andere specifieke gegevens

Beide opties bespreken we hieronder gecombineerd omdat het grotendeels een herhaling is van wat hiervóór onder de genoemde condities is opgemerkt.

Voor koppeling met BSN-gegevens zijn de beste koppelresultaten te verwachten vanwege de maximale onderscheidbaarheid van personen. Maar dit is alleen mogelijk door gegevens te versleutelen (zie koppeling met ruwe politiegegevens). Daarnaast is het zo dat als ook politie-informatie in een koppeling op basis van BSN betrokken gaat worden, dit vooralsnog alleen kan met ruwe politiegegevens omdat BSN-gegevens (vooralsnog) niet voor BRON geleverd worden.

Als het niet mogelijk of nodig is om BSN-gegevens te benutten (bijvoorbeeld als ook de huidige BRON in de koppeling meegenomen dient te worden), dan ligt het meer voor de hand om een uitvraag te doen op basis van specifieke ongevallen- en slachtoffergegevens. Omdat voor verkrijging van ernstig gewonden in principe wel een echte koppeling van gegevens het meest

wenselijke is, zou gekeken kunnen worden of uitlevering van gegevens in het huidige proces bij het CBS tot de mogelijkheden behoort of dat RAV's van mening zijn dat de gevraagde gegevens te gedetailleerd zijn en derhalve een anonimiseringslag vereisen. In dat geval is de betrokkenheid van een TTP die de persoonsgegevens versleutelt (het CBS ligt in dat geval voor de hand) noodzakelijk.

In beide gevallen kan de reeds gebruikte koppeltechniek met afstandsfunctie, en mogelijk op termijn ook een meer probabilistische koppeltechniek, worden toegepast. Daarnaast is in het geval van werkelijke aantallen een extra stap nog noodzakelijk: die van de bijschatting. Hiermee wordt geschat welk deel van de slachtoffers of ongevallen in de koppelde bestanden nog ontbreekt.

#### *Betrokken partijen*

Voor koppeling op basis van BSN: zie 'koppeling met ruwe politiegegevens' voor de betrokken partijen, waarbij LBZ-bronhouder DHD dient te worden meegenomen. Bij koppeling op basis van andere specifieke gegevens lijkt het proces meer op dat wat beschreven is bij BRON, waarbij ook hier DHD een relevante bronhouders is om afspraken mee te maken.

Bovenstaande opties zijn in *Tabel 2* samengevat.



**Tabel 2.** Koppelmogelijkheden van verschillende bronnen en daarbij benodigde koppelgegevens. PC6/PC 4 staat voor respectievelijk 'postcode compleet' (PC6) of alleen de cijfercode daaruit (PC4).

Koppelmogelijkheden	Combinatie met ambulancedata		
	BRON	Ruwe politiedata	LBZ
Koppeling op basis van BSN	Momenteel niet mogelijk vanwege het ontbreken van het BSN in BRON. Het BSN komt in de toekomst mogelijk weer wel in BRON beschikbaar; in dat geval is versleuteling nodig.	Mogelijk, wel toestemming van JenV nodig aan politie en verwerker voor levering van deze gegevens; versleuteling van BSN nodig; vooral onderzoekspotentie (o.a. verbetering BRON via andere maatschappelijke klassen en tekstherkenning); hoge koppelkwaliteit.  Vooralsnog vooral interessant voor onderzoeksdoeleinden en indirect verbetering van BRON.	Mogelijk, wel versleuteling van BSN nodig; hoge koppelkwaliteit.  Vooralsnog vooral interessant voor onderzoeksdoeleinden en verbetering van vaststelling van het aantal ernstig verkeersgewonden en kenmerken van het ongeval waarin ze betrokken waren.
Koppeling op basis van (persoons)kenmerken	Wordt reeds toegepast in koppeling met de LBZ; voor hoge koppelkwaliteit en nuttige resultaten zijn volgende gegevens nodig: <ul style="list-style-type: none"> <li>&gt; ongevalsdatum en -tijd;</li> <li>&gt; locatie van de ambulance (gps, of anders PC6 + straat+ huisnummer of desnoods PC4 + straat);</li> <li>&gt; leeftijd slachtoffer (geboortedatum of anders leeftijd + geboortedag of leeftijd + geboortemaand) en geslacht;</li> <li>&gt; vervoermiddel van slachtoffer en tegenpartij.</li> </ul> <p>Deze variant is vooralsnog het meest aan te bevelen voor de verrijking van BRON.</p>	Zie BRON.	Wordt reeds toegepast in koppeling met BRON; voor hoge koppelkwaliteit en nuttige resultaten zijn volgende gegevens nodig: <ul style="list-style-type: none"> <li>&gt; ongevalsdatum en -tijd</li> <li>&gt; BSN of leeftijd slachtoffer (geboortedatum of anders leeftijd + geboortedag of leeftijd + geboortemaand) en geslacht;</li> <li>&gt; vervoermiddel van slachtoffer en tegenpartij;</li> <li>&gt; vervoerd naar ziekenhuis;</li> <li>&gt; naam ziekenhuis.</li> </ul> <p>Deze variant is vooralsnog aan te bevelen voor de verrijking van ernstig verkeersgewonden in het bestaande proces met BRON.</p>
Combinatie op basis van open data (niet herleidbaar tot personen)	Geen koppeling mogelijk, alleen combinaties van overeenkomende clusters, vooral op basis van overeenkomst in ongevalslocatie.	Zie BRON.	Geen koppeling mogelijk wegens ontbreken persoonsgegevens.

## 4 Aanbevelingen

**Om te zorgen dat de benodigde koppelcriteria zo goed mogelijk worden meegenomen bij het verzamelen en ontsluiten van ambulancedata, is het zaak om een aantal stappen te zetten. Deze komen in dit laatste hoofdstuk aan bod.**

Het is in de eerste plaats belangrijk om met de betrokken partijen goed de doelen te bepalen waarop verdere inzet gericht zal zijn. Het ligt voor de hand hierbij in te zetten op verrijking van het verkeersongevallenbestand BRON om hiermee te voorzien in de informatiebehoefte van beleidsmakers, en daarnaast op verrijking van ernstig verkeersgewonden die tot stand komen door BRON te koppelen aan de ziekenhuisregistratie LBZ. Hierbij is het ook aan te bevelen om te kijken welke mogelijkheden benut kunnen worden om ook via andere wijzen BRON te verbeteren, bijvoorbeeld door ook ruwe politiedata erbij te betrekken. SWOV doet momenteel onderzoek naar de mogelijkheden hiervan.

Voor zowel een goede koppeling van gegevensbronnen als een bruikbaar eindresultaat (verrijkt BRON of verrijkte informatie over ernstig verkeersgewonden), zou in ieder geval aangedrongen moeten worden op uitlevering van de volgende gegevens vanuit de ambulanceregistratie:

- > ongevalsinformatie;
- > ongevalsdatum en -tijd;
- > locatie van de ambulance (gps, of anders PC6 + straat+ huisnummer of desnoods PC4 + straat);
- > slachtofferinformatie;
- > BSN of leeftijd slachtoffer (geboortedatum of anders leeftijd – geboortedag of leeftijd – geboortemaand) en geslacht;
- > vervoermiddel van slachtoffer en tegenpartij;
- > vervoerd naar ziekenhuis (voor koppeling met de LBZ);
- > naam ziekenhuis (voor koppeling met de LBZ).

Deze gegevens sluiten tevens aan bij de informatiebehoefte van beleidsmakers.

Vervolgens is het zaak om met de betrokken partijen verder te verkennen wat nodig en mogelijk is, met name ten aanzien van het detailniveau van gegevens. Het ligt hierbij voor de hand om op basis van deze notitie en de gestelde doelen afspraken te gaan maken met data-eigenaren (de ambulancevoorzieningen). Daarbij is het nodig dat zowel de ambulancevoorzieningen als verantwoordelijken van de ambulanceregistratie hierover afzonderlijk uitspraken doen en goedkeuring verlenen, maar ook dat in de behoefte wordt voorzien om dit in een centraal proces te gaan faciliteren, bijvoorbeeld via de ambulancekoepel AZN.

Hieruit volgt wat dit daadwerkelijk betekent voor overige te betrekken partijen: bijvoorbeeld welke partijen als vertrouwde derde partij (een TTP) betrokken dienen te worden; welke partijen de koppeling en andere analyses uitvoeren en welke partijen een rol spelen bij de ontsluiting van de gegevens aan eindgebruikers en wie dat precies zijn. De invulling hiervan voor zover nu bekend is voor verschillende scenario's in *Paragraaf 3.4* samengevat.

Omdat het er nu naar uitziet dat er bij de ambulancevoorzieningen een aantal voorlopers zijn die mogelijkheden voor samenwerking zien, ligt het in de rede om eerst met deze voorlopers te verkennen wat mogelijk is. Inzet zou een eerste proeflevering kunnen zijn waarbij zowel het proces kan proefdraaien maar ook op basis van de gedetailleerdere data nader in kaart kan worden gebracht hoe de verschillende databronnen zich in meer detail tot elkaar verhouden en welke inzichten dit oplevert over het ongevallen- en slachtofferbeeld in Nederland. Deze inzichten kunnen vervolgens ook weer met beleidsmakers worden gedeeld om hen goed op de hoogte te houden van de potentie van ambulancedata als aanvullende gegevensbron.

Vervolgens is het belangrijk om een plan uit te werken om het landelijke traject verder in te richten. Hierbij is een stapsgewijze benadering aan te bevelen omdat keuzes en te maken stappen sterk met elkaar samenhangen en afhankelijk zijn van de inzichten van verschillende partijen.

## Literatuur

- Bos, N.M., Decae, R.J., Bijleveld, F.D., Hermens, F., et al. (2019). *Ernstig verkeersgewonden 2018; Schatting van het aantal ernstig verkeersgewonden in 2018*. R-2019-23. SWOV, Den Haag.
- Christen, P. (2012). *Data Matching*. Springer, Berlin, Heidelberg.
- Davidse, R.J., Duijvenvoorde, K. van, Boele, M.J., Doumen, M.J.A., et al. (2014). *Fietsongevallen van 50-plussers: karakteristieken en ongevalsscenario's van enkelvoudige ongevallen en botsingen met overig langzaam verkeer; Een dieptestudie naar fietsongevallen met 50-plussers in de regio's Hollands Midden en Haaglanden*. R-2014-3A. SWOV, Den Haag
- Enamorado, T., Fifield, B. & Imai, K. (2018). *Using a probabilistic model to assist merging of large-scale administrative records*. Available at SSRN: <https://dx.doi.org/10.2139/ssrn.3214172>
- Ministerie van Infrastructuur en Waterstaat (2018). *Landelijk Actieplan Verkeersveiligheid 2019-2021; Veilig van deur tot deur*. Ministerie van Infrastructuur en Waterstaat, Den Haag.
- Ministerie van Infrastructuur en Waterstaat et al. (2018). *Veilig van deur tot deur. Het Strategisch Plan Verkeersveiligheid 2030: Een gezamenlijke visie op aanpak verkeersveiligheidsbeleid*. Ministerie van Infrastructuur en Waterstaat, Den Haag.
- Nijman, S., Klein Wolt, K., Olij, B. & Blatter, B. (2020). *Ambulancedata verkeersslachtoffers – Ontwikkeltraject landelijke ontsluiting ambulancedata voor verkeersveiligheidsbeleid*. Rapport 866, VeiligheidNL, Amsterdam.
- Olij, B. & Nijman, S. (2020). *Verkeersongevallen 2018 in Utrecht. Cijfers op basis van ambulance- en politiedata*. Rapport 852. VeiligheidNL, Amsterdam.
- Reurings, M.C.B. & Bos, N.M. (2009). *Ernstig gewonde verkeersslachtoffers in Nederland in 1993-2008; Het werkelijke aantal in ziekenhuis opgenomen verkeersslachtoffers met een MAIS van ten minste 2*. R-2009-12. SWOV, Leidschendam.
- Rijkswaterstaat (2018). *Eindrapport kwaliteitsverbetering informatieketen verkeersongevallenregistratie*. Rijkswaterstaat, Ministerie van Infrastructuur en Waterstaat, Den Haag.
- Schepers, P., Vliet, P. van, Mak, P., Aarts, L., et al. (2018). *Behoeftte aan ambulancedata en opties voor verwerking*. Discussienotitie voor Ambulancezorg Nederland. Rijkswaterstaat, Rijswijk.
- Wijlhuizen, G.J. & Bos, N. (te verschijnen). *Verkeersslachtoffers in Flevoland; Vergelijking van data uit de verkeersongevallen-, ziekenhuis- en ambulanceregistratie*. R-2020-10. SWOV, Den Haag. [Te verschijnen].

## Bijlage A Koppelvraag in de online vragenlijst

Vraag over koppeling van gegevens zoals die is opgenomen in de online vragenlijst:

Politie- en ambulancedata kunnen mogelijk aan elkaar gekoppeld worden om ervoor te zorgen dat de registratie van verkeersslachtoffers zo volledig mogelijk is. Zou deze koppeling voor u zinvol zijn en waarom?

- ja, want.....
- nee, want.....
- weet ik niet, want.....

## Bijlage B Kenmerken per bron en koppelmogelijkheden

Lijst met kenmerken per bron, de mogelijkheden die dit biedt voor koppeling van bestanden, de mate waarin een variabele onderscheidend vermogen heeft tussen slachtoffers en de beoogde einduitkomst van de koppeling.

BRON	Ruwe politiedata	LBZ	Ambulancedata	Onderscheidend vermogen	Einduitkomst relevant voor wegbeheerders
Maatschappelijke klasse 'verkeersongeval'	Maatschappelijke klasse 'verkeer' e.a.		Verkeersongevallen: type ongeval/ vervoerswijze/ vrije tekst	filtervariabele	Onderwerp: verkeersslachtoffers
Datum ongeval: jaar-maand-dag	Datum ongeval - jaar-maand-dag		Datum melding - jaar-maand-dag	Matig	Datum ongeval
Tijdstip ongeval: jaar - maand-dag-uur-minuut	Tijdstip ongeval: jaar - maand-dag-uur-minuut		Tijdstip melding: jaar - maand-dag-uur-minuut	Matig	Tijdstip ongeval
	BSN betrokkenen	BSN slachtoffer	BSN slachtoffer	Maakt unieke persoonskoppeling mogelijk	
Geboortedatum slachtoffer: jaar-maand-dag	Geboortedatum slachtoffer: jaar-maand-dag	Geboortedatum slachtoffer: jaar-maand-dag	Geboortedatum/leeftijd slachtoffer: jaar-maand-dag/leeftijd	Hoog-matig, maar afhankelijk van detailniveau	Leeftijd slachtoffer (evt. in klassen)
Ongevalselocatie: straat/kruispunt	Ongevalselocatie: straat/kruispunt		Ongevalselocatie: GPS/PC6+straat+nr/P C4+straat	Hoog-matig, maar afhankelijk van beschikbaarheid en detailniveau	Ongevalselocatie
Vervoermiddel slachtoffer (gecorrigeerd op basis van RDW)	Vervoermiddel slachtoffer	Vervoermiddel slachtoffer (grof)	Vervoermiddel slachtoffer (grof)	Matig	Vervoermiddel slachtoffer
Geslacht slachtoffer	Geslacht slachtoffer	Geslacht slachtoffer	Geslacht slachtoffer	Laag	Geslacht slachtoffer
Vervoermiddel tegenpartij (gecorrigeerd op basis van RDW)	Vervoermiddel tegenpartij	Vervoermiddel tegenpartij (grof)	Vervoermiddel tegenpartij (grof)	Matig	Vervoermiddel tegenpartij
		Gemeente ziekenhuis	Afleveradres	Matig	
		Opnametijdstip	Aflevetijdstip		
Vervoerd naar ziekenhuis Ja/Nee		Letsel slachtoffer (daaruit afgeleid ernst)	Eerste hulp geen vervoer of vervoerd naar ziekenhuis	Laag	Letselernst slachtoffer

BRON	Ruwe politiedata	LBZ	Ambulancedata	Onderscheidend vermogen	Einduitkomst relevant voor wegbeheerders
(Omstandigheden ongeval)					(Omstandigheden ongeval)
Partij- en slachtoffernummer	Partij- en slachtoffernummer		Aantal patiënten behandeld	Laag	

Hieronder is de lijst met kenmerken uit de Basisset Ambulancezorg (BSA) gegeven die relevant zijn voor verkeersveiligheid, het gewenste detailniveau van data en eventuele alternatieven in volgorde van wenselijkheid en de koppelmogelijkheden die dit biedt met verschillende bronnen. Hierin zijn ook spoedeisende hulp en locatiebronnen opgenomen, alsmede waarvoor het kenmerk relevant is.

Gegevenstype	Gewenst detailniveau	Noodzakelijk voor koppeling met:				Belangrijk voor wegbeheerders
		BRON	LBZ	SEH-bronnen	NWB/GIS	
1. Datum van rit	Dag-maand-jaar	✓	✓	✓		Wanneer vond het ongeval plaats
2. Datum en tijd van melding	Minuut – uur - dag-maand-jaar	✓ <sup>18</sup>	(✓)	(✓)		
3. Geboortedatum/leeftijd slachtoffer	In volgorde van wenselijkheid: Dag-maand- jaar Dag – jaar Maand –jaar Leeftijd - jaar	✓	✓	✓		Kenmerken slachtoffer
4. Ongevalslocatie	In volgorde van wenselijkheid: GPS zo veel mogelijk (van meldkamer of van ambulance bij stop op ongevalslocatie). PC6 + straatnaam + huisnummer (ter hoogte van) of wegnummer + hm-paal PC4 + straatnaam	✓			✓	Ongevalslocatie (belangrijke toegevoegde waarde in output voor m.n. fietsslachtoffers)
5. Vervoerswijze slachtoffer	Lijst uit BSA (AAW): Auto, Bromfiets/scooter, Bus, Fiets, Motor, Tram/Metro, Trein, Voetganger, Vrachtauto, Scootmobiel, Ebike, Speedbike, Overig  NB: dit lijstje is niet van elke RAV verkrijgbaar. Indien nodig verkrijgen door automatische tekstherkenning toepassen op opmerkingen velden. In dat geval is er mogelijk meer detailinformatie over vervoerswijzen beschikbaar.	✓	✓	✓		Kenmerken slachtoffer
6. Geslacht slachtoffer	M/V/onbekend	✓	✓	✓		Kenmerken slachtoffer



<sup>18</sup> Ligt dichterbij ongevalstijdstip dan datum van rit.



Gegevenstype	Gewenst detailniveau	Noodzakelijk voor koppeling met:				Belangrijk voor wegbeheerders
		BRON	LBZ	SEH-bronnen	NWB/GIS	
7. Vervoermiddel tegenpartij (indien bekend)	Zie 5 Evt. als minimum: Motorvoertuig betrokken in ongeval En wel/geen vervoermiddel als tegenpartij	✓				Kenmerken tegenpartij, type botsing
8. Afleveradres (indien ziekenhuis/SEH-post)	Naam ziekenhuis/SEH-post / instellingsnummer/onbekend		✓	✓		
9 Aflevertijdstip	In volgorde van wenselijkheid: Minuut–uur - dag-maand-jaar Uur - dag -maand-jaar Dag – maand - jaar		(✓) <sup>19</sup>	(✓) <sup>20</sup>		

Ad 1 en 2: deze zijn van belang om te bepalen welke regel in de te koppelen bronnen (BRON, LBZ et cetera) bij de ambulancerit horen. Er kan tijdsverschil zijn tussen het ongeval, de melding daarvan, de ambulancerit, de aankomst op de spoedeisende hulp (SEH) en opname in het ziekenhuis.

Ad 3. De geboortedatum (of desnoods leeftijd) van het slachtoffer is een zeer belangrijke koppelvariabele. Deze variabele is behoorlijk onderscheidend en wordt over het algemeen goed geregistreerd.

Ad 4. De ongevalslocatie heeft toegevoegde waarde voor koppeling met BRON maar vooral ook voor de output: we verwachten zo van met name veel fietsslachtoffers de ongevalslocatie terug te vinden. Deze gegevens ontbreken grotendeels in BRON en ziekenhuizen registreren geen locaties van het ongeval.

Ad 5. De vervoerswijze van het slachtoffer is een variabele die vooral nuttig is om na te gaan wat de kwaliteit van de gegevens is. Daarnaast is de vervoerswijze ook een belangrijke uitkomstvariabele (kenmerken van slachtoffer- en tegenpartij) voor beleidsmakers en daarom belangrijk om mee te nemen. Dit kenmerk komt in de andere bronnen ook voor en wordt vooral door de politie/RDW (BRON) naar verwachting het beste geregistreerd.

Ad 6. De variabele ‘geslacht van het slachtoffer’ wordt over het algemeen goed geregistreerd maar heeft een kleine onderscheidende waarde (er zijn maar drie mogelijkheden: man, vrouw, anders/onbekend). Toch kan deze variabele helpen bij verbetering van de koppeling met andere bronnen.

Ad 7. Zie 5.

Ad 8. Het afleveradres van het ziekenhuis is van belang bij de koppeling met een ziekenhuisopname of SEH-behandeling.

Ad 9. De dag en het tijdstip waarop de patiënt wordt afgeleverd bij de bezorglocatie, is van belang bij de koppeling met medische registraties.



<sup>19</sup> Ligt dicht bij opnametijdstip dan datum van rit.

<sup>20</sup> Ligt dicht bij binnenkomsttijdstip dan datum van rit.

## Toelichting koppelcriteria

Koppeling tussen bestanden is alleen mogelijk als verschillende databronnen vergelijkbare variabelen overeenkomstig hebben. Koppelen is vooral nuttig als deze databronnen daarnaast elkaar aanvullen en dus juist ook variabelen of gevallen bevatten die wel of beter in de ene maar niet of minder goed in de andere bron voorkomen, of meer relevante informatie binnen een variabele bevatten waardoor een beeld ontstaat van de overlap en (in)completeheid van de bronnen. Bij het koppelen van bronnen is daarnaast de kwaliteit van de data en met name de koppelvariabelen belangrijk (bij een lage kwaliteit is de betrouwbaarheid van de koppeling laag) en ook wat onderscheidend vermogen ervan is (bij weinig variatie in de waarden, kan lastiger onderscheid worden gemaakt welke regels daadwerkelijk bij elkaar horen en is in het slechtste geval alleen sprake van een combinatie van bestanden).

## In- en exclusiecriteria

Naast bovengenoemde data die nodig zijn om optimaal te kunnen koppelen, is het ook belangrijk te weten of het om een verkeersongeval ging of niet. Dit kan op drie manieren:

- Type ongeval: er wordt aangegeven of het om een verkeersongeval gaat of niet.
- Vermelding van vervoermiddel slachtoffer en eventueel ook de tegenpartij: hieruit kan indirect afgeleid worden dat het om een verkeersongeval ging, ook wanneer het type ongeval onbekend is of niet ingevuld.
- Toelichting (vrije tekst): uit de vrije tekstvelden kan opgemaakt worden of het om een verkeersongeval ging door middel van tekstherkenning.

De verwachting is dat in ieder geval van een deel van de regio's de manieren 1 en 2 een te onvolledig bestand opleveren. Aanvulling op basis van informatie uit vrije tekstvelden is in die regio dan noodzakelijk. Dit kan door automatisch te laten zoeken naar tekst die verwijst naar op de openbare weg gebruikte vervoermiddelen of andere bewoording waaruit een verkeersongeval kan worden afgeleid.

## Koppelkwaliteit

De koppelkwaliteit hangt af van een aantal kenmerken, te weten:

- De aan- of afwezigheid van koppelvariabelen (zie lijst).
- De mate van detail van de aanwezige koppelvariabelen (is de meest gewenste optie aanwezig, of de minder wenselijke grovere opties? Bijvoorbeeld tijdstip in uren of in minuten, zie lijst).
- Het onderscheidende vermogen van de variabele (een variabele met meer mogelijke waarden (bijvoorbeeld geboortedatum) heeft een groter onderscheidend vermogen dan een variabele met minder waarden (bijvoorbeeld geslacht)).
- De kwaliteit van de variabele (hoe goed geeft het de werkelijkheid weer; dit hangt bijvoorbeeld mede af van de wijze van registratie (automatisch of handmatig) en eventuele benodigde beoordelingsdeskundigheid of - noodzaak (letsel of naar welk ziekenhuis vervoerd zal een hogere prioriteit hebben en beter worden geregistreerd dan bijvoorbeeld de vervoerswijze, zeker als dat nieuwe of specifieke typen betreft).

Onderstaande tabel geeft het aantal mogelijke unieke waarden aan per variabele (binnen een jaar). Door het aantal unieke waarden per variabele met elkaar te vermenigvuldigen, kunnen we het onderscheidende vermogen van de gehele set aan variabelen bepalen. Combineren we dat met het aantal ambulanceritten dat beschikbaar is, dan kunnen we een beeld krijgen hoe onderscheidend de set van data in totaal zal zijn (en hoe groot de kans op unieke combinaties; dit is ook een maat voor de kwaliteit van de koppeling).

Gegevenstype	Detailniveau	Onderscheidend vermogen (aantal mogelijke waarden)
1. Datum van rit	Dag-maand-jaar	365
2. Datum en tijd van melding	Minuut – uur - dag-maand-jaar	$60*24*365 = 525.600$
3. Geboortedatum/ leeftijd slachtoffer	1. Dag-maand- jaar	365
	2. Dag – jaar	31
	3. Maand –jaar	12
	4. Leeftijd – jaar	Ca. 100
4. Ongevalslocatie	1. GPS zo veel mogelijk (van meldkamer of van ambulance bij stop op ongevalslocatie).	Ca 33.893 <sup>21</sup>
	2. PC6 + straatnaam + huisnummer (ter hoogte van) of wegnummer + hm-paal	Ca. 457.000 <sup>22</sup>
	3. PC4 + straatnaam	Ca. 4.100 <sup>23</sup>
5. Vervoerswijze slachtoffer	1. Auto, Bromfiets/scooter, Bus, Fiets, Motor, Tram/Metro, Trein, Voetganger, Vrachtauto, Scootmobiel, Ebike, Speedbike, Overig	13
	2. Informatie op basis van tekstherkenning	>13
6. Geslacht slachtoffer	M/V	2
7. Vervoermiddel tegenpartij (indien bekend)	1. Auto, Bromfiets/scooter, Bus, Fiets, Motor, Tram/Metro, Trein, Voetganger, Vrachtauto, Scootmobiel, Ebike, Speedbike, Overig	13
	2. Informatie op basis van tekstherkenning	>13
	3. Motorvoertuig betrokken in ongeval	(wel/niet) = 2
	4. Wel/geen vervoermiddel als tegenpartij	(wel/niet) = 2
8. Afleveradres (indien ziekenhuis/SEH-post)	Naam ziekenhuis/SEH-post / instellingsnummer	116 <sup>24</sup>
9 Aflevertijdstip	1. Minuut – uur – dag – maand - jaar	$60*24*365 = 525.600$
	2. Uur - dag -maand - jaar	$24*365 = 8.760$

Het maximale onderscheidende vermogen voor koppeling met BRON kan in principe worden bepaald door na te gaan welke van bovenstaande kenmerken in de ambulancedata aanwezig zijn en in welk detailniveau. Het aantal mogelijke waarden van iedere variabele wordt vervolgens vermenigvuldigd met het aantal mogelijke waarden in andere variabelen. Dit wordt vergeleken met de totale omvang van de bestanden die gekoppeld worden.



21 <https://www.wegenwiki.nl/Nederland>; blokken van 1 x 1 kilometer in Nederland (= de landoppervlakte in km<sup>2</sup>), dit is wat minder dan de weglengte in Nederland (ca. 140.000 km), omdat veel straten parallel lopen.

22 [Postcodeinfo.nl](https://www.postcodeinfo.nl)

23 <https://allecijfers.nl/definities/open-data-bronnen-informatie-per-postcode/>

24 <https://www.volksgezondheidenzorg.info/onderwerp/ziekenhuiszorg/regionaal-internationaal/locaties#!node-algemene-en-academische-ziekenhuizen>, buitenpoli niet meegeteld.

Daarbij moet wel rekening gehouden worden met een aantal zaken die de berekening lastiger maken zoals:

- Kleine toelaatbare verschillen tussen de registraties. Wanneer bijvoorbeeld het uur/ minuut van een ongeval, melding of opname niet exact overeenkomt, betekent dat niet dat regels in verschillende bestanden die hierop betrekking hebben niet bij elkaar horen. Dit hangt ook weer samen met andere kenmerken (bijvoorbeeld: als er lichter letsel is, is de kans groter dat er zowel tussen ongevals- en opnametijd en mogelijk ook tussen ongevals- en opnamelocatie grotere verschillen zitten).
- Onderscheidbaarheid binnen groepen afhankelijk van groepsgrootte: bepaalde leeftijdsgroepen komen bijvoorbeeld veel vaker voor in ongevallen dan andere. De piek ligt bij de leeftijden 16 – 18 jaar. Binnen dergelijke grote groepen is een groter onderscheidend vermogen nodig om individuele slachtoffers binnen een bestand te onderscheiden dan voor kleinere groepen.

## Bijlage C Verwerking van BSN

Voor koppeldoelinden is een bestand waarbij individuen uniek te herkennen zijn, de meest ideale situatie. Een van de geschiktste persoonsidentificatienummers is het burgerservice-nummer (BSN). Alle inwoners van Nederland (ingezetenen) hebben een BSN en zouden op basis daarvan identificeerbaar moeten zijn. Alleen niet-ingezetenen kunnen hiermee niet geïdentificeerd worden. Hiervoor moeten andere methoden worden gezocht.

Vanuit de bescherming van persoonsgegevens (Algemene verordening gegevensbescherming AVG en Uitvoeringswet Algemene verordening gegevensbescherming UAVG) zijn er echter voorwaarden gesteld aan de verwerking van BSN. Het BSN is weliswaar geen 'bijzonder persoonsgegeven', maar de verwerking van dit nummer is voorbehouden aan specifieke organisaties en situaties, te weten:

- Organisaties die bij wet een taak hebben om personen uniek te kunnen identificeren, zoals overheidsorganisaties, artsen, werkgevers en banken.
- Echter, zij mogen dit alleen gebruiken als het BSN voor die taak noodzakelijk is. Dit wil concreet zeggen dat het moet gaan om een wettelijk vereiste situatie waarbij het noodzakelijk is zeker te weten om welke persoon het gaat.

Voor koppeling van ongevalgegevens betekent dit dat:

- Organisaties die dit nummer willen gebruiken, hiervoor een wettelijke taak dienen te hebben of te krijgen waarbij benutting van het BSN expliciet wordt opgenomen. Het gaat hierbij overigens niet zozeer om het identificeren van personen als wel om goed onderscheid te kunnen maken tussen personen. Daarvoor is een afgeleide van het BSN ook geschikt.
- Organisaties die van dit nummer gebruik willen maken voor koppeling van gegevens, maar hiertoe (nog) geen wettelijke taak hebben, dit nummer alleen indirect kunnen gebruiken door middel van versleuteling. Dit is voor het doel van de koppeling overigens voldoende.
- Overigens is bovenstaande alleen zinvol als in alle te koppelen bestanden het BSN aanwezig is. Voor BRON is dat bijvoorbeeld momenteel niet het geval.

## Bijlage D Vereisten bij de verwerking van patiëntgegevens

Bij verwerking van ambulancegegevens (en dus ook bij koppeling van deze gegevens aan andere bestanden) hebben we te maken met gegevens die van medische aard kunnen zijn. Dit is zeker het geval als er informatie uit vrije tekstvelden wordt gewerkt (hier kunnen immers medische details in staan). Minder duidelijk is dit overigens als slechts een selectie van vooraf bekende kenmerken die niet specifiek medisch van aard zijn worden verwerkt. Wel kan in dat geval afgeleid worden dat de betreffende persoon verwondingen had of ten minste vermoedens daarvan, en dat kan als medisch gegeven worden beschouwd. Hoe dan ook hebben we bij de verwerking van dergelijke gegevens te maken met de AVG (Algemene verordening gegevensbescherming) en UAVG (Uitvoeringswet Algemene verordening gegevensbescherming), omdat er ten minste tot personen herleidbare gegevens in het spel zijn. Daarnaast betreft het gegevens die vallen onder de Wet op de geneeskundige behandelovereenkomst (WGBO), waarvoor speciale afspraken gelden.

Hieronder zetten we uiteen wat dit betekent voor de verwerking van ambulancegegevens en aan welke voorwaarden bij de verwerking (ook bij koppeling) van deze gegevens moet worden voldaan en onder welke condities. Hierbij baseren we ons grotendeels op de *Handreiking ontsluiting patiëntgegevens voor wetenschappelijk onderzoek*<sup>25</sup>.

### Verwerkingsmogelijkheden volgens de AVG en UAVG

Volgens de AVG zijn medische gegevens 'bijzondere persoonsgegevens' waarvoor speciale verwerkingsafspraken gelden<sup>26</sup>: verwerking hiervan is in principe niet toegestaan, tenzij aan bepaalde voorwaarden wordt voldaan. We benoemen hieronder de relevante mogelijkheden:

- Algemeen belang: het gaat hierbij om een wettelijke taak van de organisatie (kan ook een ministeriële beschikking zijn), waarbij het noodzakelijk is om de betreffende persoonsgegevens te verwerken in het kader van het algemeen belang (in dit geval: preventie van verkeersongevallen).
- Algemeen belang: verwerking is noodzakelijk voor archiveringsdoeleinden, wetenschappelijk of historisch onderzoek of statistiek en hierbij is sprake van een wettelijke taak.

Verwerking is daarnaast ook toegestaan indien de betrokkene hiertoe expliciet toestemming heeft gegeven of de gegevens zelf openbaar heeft gemaakt.



<sup>25</sup> <https://www.rivm.nl/documenten/handreiking-ontsluiting-patientgegevens-voor-wetenschappelijk-onderzoek>

<sup>26</sup> <https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/algemene-informatie-avg/mag-u-persoonsgegevens-verwerken#wat-verstaat-de-avg-onder-bijzondere-persoonsgegevens-6339>; geraadpleegd op 11 jun 2020.

## Verwerkingsmogelijkheden volgens de WGBO

Volgens de WGBO zijn verwerkingen van medische gegevens alleen toegestaan voor organisaties die bij het proces van afhandeling betrokken zijn en hierbij dus een specifieke rol hebben. Alleen onder strikte voorwaarden zijn deze gegevens met derden te delen. Omdat deze eisen specifiek zijn dan de AVG, gaat de WGBO vóór de AVG en hebben we vooral met de WGBO rekening te houden.

Indien medische gegevens worden gedeeld met onderzoekers, dient ten minste aan de volgende voorwaarden te worden voldaan:

- Dataveiligheid: alleen specifiek daartoe aangewezen onderzoekers mogen de data verwerken en hiertoe dienen zowel technische als organisatorische maatregelen te worden getroffen, conform ISO27001-standaard.
- GEB of DPIA: Voorafgaand aan het opzetten van een onderzoek of project, dient een gegevensbeschermingseffectbeoordeling (GEB of: data protection impact assessment, DPIA) te worden uitgevoerd.
- Duurzame opslag: onderzoeksgegevens dienen opvraagbaar te zijn volgens de principes van FAIR (= Findable, Accessible, Interoperable and Reusable).

Daarnaast wordt ook nog de toetsing door een toetsingscommissie genoemd, maar dit is niet nodig voor retrospectief onderzoek. De inzage- en rectificatierechten van patiënten kan in geval van wetenschappelijk onderzoek onder bepaalde voorwaarden worden ingeperkt.

Bij de verwerking van patiëntgegevens zijn er om te beginnen drie opties die te maken hebben met goedkeuring door de patiënt:

1. Expliciete toestemming van de patiënt (voorafgaand of achteraf aan de gegevensverzameling).
2. Het principe van 'geen bezwaar' is van kracht.
3. Gegevens worden geanonimiseerd.

### Toestemming van patiënten

Bij de verwerking van ambulancegegevens is toestemming van patiënten in principe niet aan de orde: verwerkers hebben in principe geen contactgegevens van de betrokkenen, kennen de betrokkenen vooraf niet en het kost buitensporig veel tijd en middelen en schendt juist meer de vertrouwelijkheid indien vervoerde slachtoffers om toestemming gevraagd zou worden. De optie 2 en 3 zijn dus relevant om te (ver)kennen.

### Eisen bij 'geen bezwaar'

Bij de optie 'geen bezwaar' gelden de volgende eisen:

- Het onderzoek dient een algemeen belang en de resultaten dienen publiek beschikbaar te komen.
- Zonder deze gegevens kan het onderzoek niet worden uitgevoerd (anonimiseren is onmogelijk).
- Er wordt voldaan aan de hiervóór genoemde voorwaarden van de AVG en de WGBO (zoals grondslagen, dataveiligheid, uitvoering van een DPIA en FAIR dataopslag).
- In het dossier van de betrokkenen wordt vermeld dat hun gegevens voor onderzoek gebruikt zijn. De zorgverlener heeft in dat geval aan de patiënt gemeld dat zijn gegevens voor wetenschappelijk onderzoek is gebruikt.
- Er gelden extra eisen aan het systeemontwerp voor garanties van dataveiligheid ('privacy by design plus'). Gegevens mogen in dit geval alleen verwerkt en ontsloten worden met gegevens die met moeite herleidbaar zijn.



De eerste drie van deze eisen lijken bij koppeling van ambulancegegevens aan de orde; de vierde eis van de ‘geen bezwaar’-optie is naar verwachting een aandachtspunt voor RAV’s en GGD’s. Extra beveiligingswaarborgen zouden gerealiseerd kunnen worden door verwerkers te laten werken volgens de informatiebeveiligingsnorm voor de zorg NEN7510. Daarnaast zouden gegevens versleuteld (gepseudonimiseerd) kunnen worden.

### **Anonimisering van gegevens**

De derde optie om medische gegevens te mogen verwerken, is als de gegevens geanonimiseerd worden. Volgens de AVG is er in dat geval geen sprake meer van persoonsgegevens en vallen de gegevens niet meer onder deze wet. Wat ‘geanonimiseerde gegevens’ zijn, is echter nog niet zo gemakkelijk uit te maken. Gegevens kunnen door een combinatie van kenmerken van persoonsgegevens immers ook indirect tot personen herleidbaar blijven, ook als er direct naar personen verwijzende kenmerken zijn uitgehaald. Als richtlijn hierbij geldt dat een slimme data-analist de gegevens niet naar een persoon zou kunnen herleiden. Dat is met name cruciaal voor de zeldzame gevallen in een databestand die het gemakkelijkst op basis van hun kenmerken herleidbaar zouden zijn naar personen (denk aan gegevens uit dun bevolkte gebieden en kenmerken van slachtoffers die weinig voorkomen zoals een weinig voorkomende leeftijd of vervoerswijze). Herleidbaarheid kan worden voorkomen door data minder specifiek te maken (bijvoorbeeld geen geboortedatum, maar alleen leeftijd of leeftijdscategorie geven en zeldzame vervoermiddelen samenvoegen in een categorie ‘overig’).

Anonimisering van gegevens mag alleen gebeuren door de volgende instanties en onder de volgende voorwaarden:

- Altijd door rechtstreeks betrokkenen bij de behandeling (zie algemene eis van de WGBO).
- Altijd door een verwerker die al een rol heeft in het primaire proces van hulp aan de patiënt en de afhandeling daarvan. In dat geval dient er wel toestemming van de verwerkersverantwoordelijke te zijn en een geldige verwerkersovereenkomst.
- De anonimisering gebeurt door een ingehuurd professionele organisatie: een vertrouwde tussenpersoon of Trusted Third Party (TTP). Een dergelijke organisatie kan bijvoorbeeld het CBS zijn (als het om gegevens gaat die niet door het CBS zelf zijn verzameld) of de Stichting ZorgTTP (die ook reeds veel anonimiseringswerk doet voor zorgpartijen zoals RAV’s en DHD, de leverancier van de LBZ).

Er wordt in de handreiking aangeraden om een Data Transfer Agreement (DTA) op te stellen in een dergelijke verwerking, ook al is dit niet verplicht. Hiermee wordt geregeld dat de data alleen voor de beoogde doeleinden worden verwerkt.

Hoe anonimisering bij een TTP in zijn werk gaat, is beschreven in *Bijlage E*.

## Bijlage E Anonimisering van gegevens via een TTP

Hieronder schetsen we de algemene stappen die bij anonimisering van gegevens worden doorlopen, evenals nadere informatie die is verkregen in een gesprek met Michiel Vlastuin en Thomas Waslijah van ZorgTTP op 24 juni 2020.

### Anonimiseringsstappen

Bij de anonimisering van gegevens via een vertrouwde derde partij (TTP), vindt het volgende standaardproces plaats in drie verschillende softwaremodules:

1. Stap bij de RAV: direct tot personen herleidbare gegevens worden bij de bron (dus bij de RAV of GGD) door software (de privacy- en verzendmodule) reeds versleuteld tot een pre-pseudoniem. Dit pre-pseudoniem wordt gescheiden van het overige datadeel en ook dit overige datadeel wordt versleuteld. Beide bestanden worden vervolgens naar de TTP gestuurd.
2. Stap bij de TTP: bij de TTP wordt het pre-pseudoniem omgezet in een uiteindelijk pseudoniem (pseudonimiseringsmodule). De TTP stuurt vervolgens zowel het pseudoniem als de versleutelde overige data naar de onderzoeker.
3. Stap bij de onderzoeker: de onderzoeker kan de overige data ontsleutelen met de verkregen sleutel en deze gebruiken voor de overeengekomen doeleinden (data- en recievemodule DRM). Koppeling van gegevens met andere bronnen gebeurt op basis van het pseudoniem. Om te zorgen dat in beide bronnen identieke personen eenzelfde pseudoniem krijgen, is het in deze werkwijze van belang dat alle te koppelen bestanden aan hetzelfde pseudonimiseringsproces worden onderworpen.

Strikt gesproken is pseudonimiseren niet in alle gevallen hetzelfde als anonimiseren. Dit is namelijk alleen het geval als de pseudonimisering niet meer ongedaan gemaakt kan worden: gegevens die via éénweg-pseudonimisering versleuteld zijn, kunnen nooit meer worden terugherleid naar de oorspronkelijke persoonsgegevens. Bij tweeweg-pseudonimisering is terugherleiding wel mogelijk door het omgekeerde versleutelingsproces te bewandelen.

### Specificaties van data

Een eerste verkennend gesprek met medewerkers van ZorgTTP leert daarnaast het volgende:

- Het is zowel mogelijk om op basis van BSN te pseudonimiseren als op basis van een combinatie van andere persoonskenmerken (meestal geboortedatum, geslacht, naam en postcode), bijvoorbeeld voor die personen waarvoor geen BSN voorhanden is. Deze pseudonimisering is echter wel minder betrouwbaar.
- Ook bij het pseudonimiseren van persoonsgegevens zoals geboortedatum, is het mogelijk om gecategoriseerde hiervan afgeleide waarden aan onderzoekers ter beschikking te stellen (bijvoorbeeld: geboortedatum wordt niet doorgeleverd, maar wel de leeftijd of leeftijdsklasse van de betrokkene).
- Welk detailniveau acceptabel is om door te leveren voor onderzoeksdoeleinden, wordt bepaald door de Functionaris Gegevensbescherming van de bronhouder (RAV of GGD). Daar moeten dus eerst afspraken mee worden gemaakt; er gelden geen algemene uitgangspunten

of wetsinterpretaties van wat acceptabel is door de TTP. In het algemeen zal gezocht worden naar het minst onthullende niveau van data dat én noodzakelijk is voor het onderzoek én acceptabel wordt geacht vanuit interpretatie van de wetgeving. Wel is duidelijk dat een BSN niet als zodanig mag worden doorgeleverd (zie ook *Bijlage C*).

## Organisatorische zaken

- Na overeenstemming dat en hoe de data via de TTP aan onderzoekers zullen worden geleverd, maakt de TTP eerst een functioneel ontwerp van welke bestanden worden verwerkt en welke controles plaatsvinden (vooral opsporen van datafouten). Op basis daarvan wordt de software ontwikkeld. Dit neemt circa zes weken in beslag, maar dat is afhankelijk van het aantal partijen dat betrokken is en hun bestandsspecificaties.
- Iedere RAV die deelneemt aan dit proces dient akkoord te zijn met de bestandsspecificatie of er dient eventueel per RAV met specifieke bestandsspecificaties gewerkt te worden. De verwachting is echter dat RAV's zich aan een landelijk ontwikkeld protocol zullen willen houden (zie ook *Bijlage F*). Als dit allemaal rond is, dienen uiteraard ook verwerkerovereenkomsten gesloten te worden. De TTP heeft hiervoor sjablonen beschikbaar.
- Daarnaast dienen afspraken goed te worden vastgelegd, net als uit te wisselen databronnen en de juridische grondslagen waaronder ze uitgewisseld worden.
- Ook dient er een DPIA uitgevoerd te worden (data protection impact assessment of gegevensbeschermingseffectbeoordeling, GEB).
- De kosten zijn op dit moment nog niet gespecificeerd maar bestaan uit de volgende posten:
  - ontwikkelkosten;
  - verwerkerskosten (onder andere het aantal verschillende formats en aantal regels in deze bestanden dat verwerkt dient te worden, alsmede ondersteuning van de servicedesk);
- contractkosten (aantal verwerkerovereenkomsten dat moet worden gesloten).

Zodra alles rond is duur het drie weken voordat de verwerking gestart kan worden. Op de frequentie van levering zit geen beperking.

## Bijlage F Beschikbare data vanuit ambulanceregio's: een casus

De GGD Flevoland is een van de voorlopers als het gaat om de benutting van ambulancegegevens voor verkeersveiligheidsanalyses. Zij hebben deze data zelf op een eigen dataportaal met aantrekkelijke gebruikersinterface gezet<sup>27</sup> en daarnaast ook op de rijkspagina voor open data beschikbaar gesteld<sup>28</sup>.

### Open data

Wat er mogelijk is aan koppeling met deze open data en BRON, is reeds in *Paragraaf 2.3.2* besproken. Data in deze open databron bevat de volgende variabelen die door de GGD/RAV worden beschouwd als niet meer tot personen herleidbaar:

- > ongevalsnummer;
- > ongevalsjaar;
- > vervoermiddel van slachtoffer en tegenpartij;
- > aantal slachtoffers in het ongeval;
- > gemeentenaam;
- > gps-locatie van de ambulance (in long-lat-coördinaten; dit is bij benadering de ongevalslocatie);
- > kruispuntnummer (gebaseerd op de kruispunten in BRON, tot op 20 meter nauwkeurig);
- > wegvaknummer (gebaseerd op het Nationaal Wegenbestand NWB, tot op 20 meter nauwkeurig).

De data dekken momenteel de jaren 2013 t/m 2019.

### Onderzoeksdata: een grofmazig koppelonderzoek

In opdracht van de provincie Flevoland zijn daarnaast ook gedetailleerdere data aan SWOV beschikbaar gesteld ten behoeve van een onderzoek om het aantal ernstig verkeersgewonden in Flevoland te kunnen vaststellen. Hierbij dient wel opgemerkt te worden dat de variabelen en het detailniveau van de data samenhangen met de onderzoeksvragen die uitgezocht dienden te worden. In ieder geval waren hierbij de volgende *extra* variabelen beschikbaar om ambulancedata te kunnen koppelen met BRON en de LBZ:

- > leeftijdsklasse van de patiënt (in groepen van in principe tien jaar);
- > of de patiënt vervoerd is en waar naartoe (behandeld ter plaatse of vervoerd naar ziekenhuis);
- > ziekenhuisnaam (naar welk ziekenhuis de patiënt is vervoerd).



<sup>27</sup> Zie <https://www.eengezonderflevoland.nl/verkeersongevallen-flevoland-almere-lelystad>

<sup>28</sup> Zie <https://data.overheid.nl/dataset/verkeersongevallen-met-ambulancEZorg-in-flevoland>

In dit onderzoek is niet een-op-een gekoppeld met BRON of de LBZ. Was dat wel de bedoeling geweest, dan zouden ook de geboortemaand of geboortedag samen met de leeftijd in jaren een minimumvereiste zijn geweest om tot een kwalitatief redelijke koppeling te komen.

In een onderzoek van VeiligheidNL (Olij & Nijman, 2020), waarbij BRON-data aan data van de Regionale Ambulance Voorziening Utrecht (RAVU) zijn gekoppeld, bleken de volgende gegevens beschikbaar te zijn:

- › leeftijdsklasse van slachtoffer (in groepen van in principe tien jaar);
- › geslacht van slachtoffer;
- › gemeente van het ongeval (van de ambulancegegevens was daarnaast de PC4 en van sommige regels de PC6 + huisnummer als locatie beschikbaar, maar deze is in de koppeling niet gebruikt);
- › datum en tijdstip van het ongeval.

Ook in dit onderzoek is geen een-op-een-koppeling gerealiseerd en is een eenvoudige deterministische koppeling uitgevoerd waarbij alleen op tijdstip ongeval tussen BRON en de ambulancedata enig tijdsverschil is toegelaten.

## Overwegingen bij gegevensverwerking ambulancegegevens Flevoland

Daarnaast is uit eerdere gesprekken en mailconversatie het volgende naar voren gekomen ten aanzien van de afwegingen die de GGD en RAV Flevoland maken bij verwerken van ambulancegegevens:

Om tot de open dataset te komen heeft de RAV Flevoland een DPIA (data protection impact assessment of gegevensbeschermingseffectbeoordeling, GEB) uitgevoerd en is op basis daarvan tot de huidige uitgeleverde data gekomen. Hierover worden patiënten overigens niet geïnformeerd. Er is sprake van de anonimiseringsoptie van de WGBO (zie *Bijlage D*).

- › De RAV Flevoland geeft aan dat de uitgevoerde DPIA niet in andere uitleveringen van data voorziet dan de open data. Zij beschouwen dit ook niet als hun primaire taak en zien landelijke uitlevering dan ook als een vraag die via koepelorganisatie AZN centraal uitgezocht zou moeten worden.
- › Bij het proces om tot de open data te komen verwerkt de GGD (die dit proces uitvoert) zelf ook geen directe persoonsgegevens zoals NAW-gegevens en geboortedatum. In dat geval wordt gesproken van ‘herleidbare gegevens’. Het is op dit moment nog niet expliciet duidelijk wat precies onder wel of niet herleidbaar wordt beschouwd, maar indirect zou de datalevering aan SWOV kunnen worden gezien als een variant van ‘niet herleidbare’ informatie.
- › Om te bepalen of een ambulancerit een verkeersongeval betrof, wordt in de data (georganiseerd volgens het model ‘ambite’) van de RAV Flevoland eerst geselecteerd op een variabele die ‘mechanisme’ heet, en daarbinnen de waarde ‘verkeer’. Dit betreft een variabele die door het ambulancepersoneel wordt ingevuld. Dit wordt overigens vanuit de module aangevuld met informatie uit de tekstvelden. Het is op dit moment niet precies bekend hoe dit in zijn werk gaat.
- › Voor de verwerking van ambulancegegevens door onderzoeksinstanties geeft de GGD aan dat er waarschijnlijk meer mogelijk is, mits hiervoor de juiste grondslagen aanwezig zijn.

## Bijlage G Schatting van overlap tussen registraties

Om voor het aantal ernstig verkeersgewonden de overlap tussen bronnen te bepalen (zie *Afbeelding 1b*), hebben we het volgende gedaan:

- We hebben het gemiddelde genomen van de werkelijke aantallen over de periode 2016-2019 (21.500 ernstig verkeersgewonden).
- Vervolgens hebben we gezien hoeveel van deze werkelijke aantallen in de verschillende bronnen teruggevonden worden:
  - BRON: 6.500 slachtoffers kunnen worden gekoppeld aan ernstig verkeersgewonden in de LBZ (30% van het totaal). De overige BRON-slachtoffers zijn licht gewond.
  - LBZ: 19.000 ernstig verkeersgewonden (88% van het totaal; op basis van Bos et al., 2019; het gaat hier om slachtoffers die als ernstig verkeersgewonden zijn geregistreerd; een deel van de gewonden die uiteindelijk ook ernstig verkeersgewonden blijken te zijn, zitten soms onder andere oorzaken in het bestand, of moest worden bijgeschat).
  - Ambulancebestand: 17.600 ernstig verkeersgewonden (afgerond op honderdtallen; 82%; 80% geldt voor alle opnamen). Deze schatting is gebaseerd op basis van de LIS-bestanden van 2009-2012. Zie bijvoorbeeld *Tabel 3.1* uit Wijlhuizen & Bos, te verschijnen.
- Vervolgens is de overlap tussen bestanden als volgt bepaald:
  - Vrijwel alle BRON-regels vinden we ook terug in de LBZ (dus:  $6.500/19.000 = 34\%$  van de LBZ heeft ook een BRON-registratie).
  - Ambulancedata in de LBZ: de overlap is hier bepaald door het product te nemen van de aandelen van elk van de bestanden in het totale aantal ernstig verkeersgewonden (respectievelijk 82% ambulance en 88% LBZ). Vervolgens is op basis van een logische combinatie van de gegevens die alles bij elkaar niet meer ernstig verkeersgewonden mogen bevatten dan het totaal, geschat dat er circa 25% extra overlap zal zijn tussen ambulancedata en de LBZ. Meer zekerheid over het werkelijke aandeel verkrijgen we pas bij een daadwerkelijke koppeling. Daarvoor ontbreken tot nog toe geschikte data. Als we kijken naar ernstig verkeersgewonden, schatten we dat 90% van de ambulanceregistraties kan worden teruggevonden in de LBZ. Circa 10% van de ambulanceregistraties is naar verwachting niet in de LBZ terug te vinden als ernstig verkeersgewonde, maar als een ernstig gewonde die later alsnog aan verkeer gerelateerd kan worden.
  - BRON in ambulancedata: ook hier is bovenstaande productregel toegepast (respectievelijk 30% en 83%) en is geschat dat ook hier circa 25% extra overlap is tussen bestanden. Dit zou betekenen dat circa 31% van de ambulanceregistraties die betrekking hebben op ernstig verkeersgewonden, ook een gekoppelde BRON-registratie zou hebben. Relateren we dit aan het aantal BRON-registraties van het aantal ernstig verkeersgewonden, dan komen we tot de conclusie dat 2.000 (= circa 10%) BRON-registraties alleen in de LBZ worden teruggevonden, en dat 4.500 (= circa 20% van het totaal) BRON-registraties zowel in de LBZ als in ambulancegegevens worden teruggevonden.
  - Rest nog de ambulanceregistraties die wel in de LBZ maar niet in BRON voorkomen. Deze worden geschat door de BRON-overlap en de niet-overlappende ambulanceregistraties af te trekken van het aantal ambulanceregistraties die we als ernstig verkeersgewonde verwachte terug te vinden. Dit leidt tot 11.000 gevallen per jaar (circa 50% van het totaal).
  - De schatting van het aandeel LBZ-registraties zonder overlap met BRON of ambulancegegevens, wordt geschat door de aantallen registraties die wél overlappen, af

te trekken van het totaal aan LBZ-registraties die als ernstig verkeersgewonde herkenbaar zijn. Dit gaat om circa 1.500 slachtoffers (circa 7% van het totaal).

- Alles met elkaar combinerend en relaterend aan het aantal ernstig verkeersgewonden, schatten we dat nog circa 400 ernstig verkeersgewonden (circa 2%) in geen van de registraties teruggevonden zal worden. Dit aantal volgt uit de bijschattingsprocedure.



## Bijlage H Toelichting op de koppelmogelijkheden van open data

In *Afbeeldingen 3 en 4* zagen we dat er op ongevalsniveau veel 'één-op-veel'-combinaties werden gevonden bij het combineren van BRON-gegevens met open ambulancegegevens voor de provincie Flevoland. Hieronder lichten we toe welke stappen gezet zijn bij het verkrijgen van dit resultaat. Deze geven aan welke handelingen nodig kunnen zijn bij het koppelen van open data.

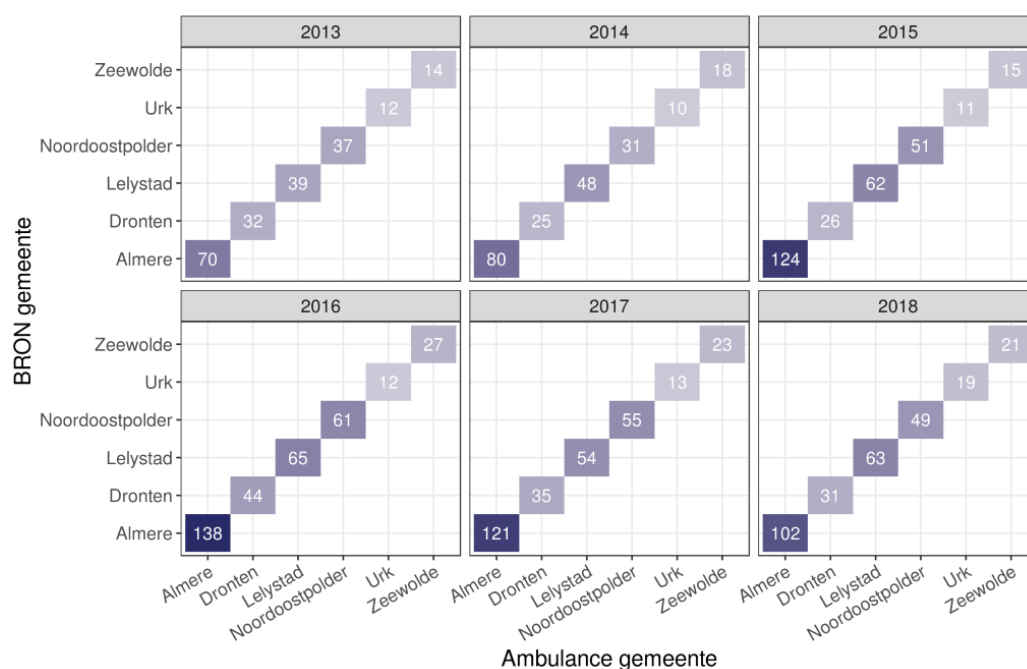
- Na het inladen van de bestanden zijn de variabelen zodanig hernoemd dat beide bestanden gebruikmaken van de voor de koppeling gebruikte variabelen 'Jaar', 'Provincie', 'Gemeente', en locatie in lengte- en breedte.
- Elk incident in BRON en in de ambulanceregistratie kreeg vervolgens een eigen ID toegewezen (respectievelijk 'BRON ID', 'ambulance ID').
- BRON maakt gebruik van rijksdriehoekcoördinaten, de ambulanceregistratie van gps. Om tot een bruikbare koppeling te komen moet van hetzelfde coördinatenstelsel worden uitgegaan en moeten de coördinaten van ene bron worden omgezet in die van de andere.
- Voor de deterministische koppeling zijn de twee bestanden vervolgens gecombineerd op de variabele 'locatie' met behulp van een techniek die bekend staat als 'fuzzy join'. Daarbij is een maximale afstand van ca. 180 m toegestaan, waardoor toegelaten wordt dat een ongeval op een iets andere plaats gebeurd kan zijn dan de plaats waar de ambulance is neergezet en haar coördinaten heeft opgeslagen.
- Vervolgens zijn alleen combinaties toegestaan waarbij ook de gemeente en het jaar overeenstemmen.
- Tot slot is per ambulanceregistratie geteld hoeveel BRON-registraties met dezelfde kenmerken konden worden gevonden, en omgekeerd. De vijftig registraties met de hoogste frequenties zijn getoond in *Afbeelding 4*.

### Aanzet tot probabilistisch koppelen

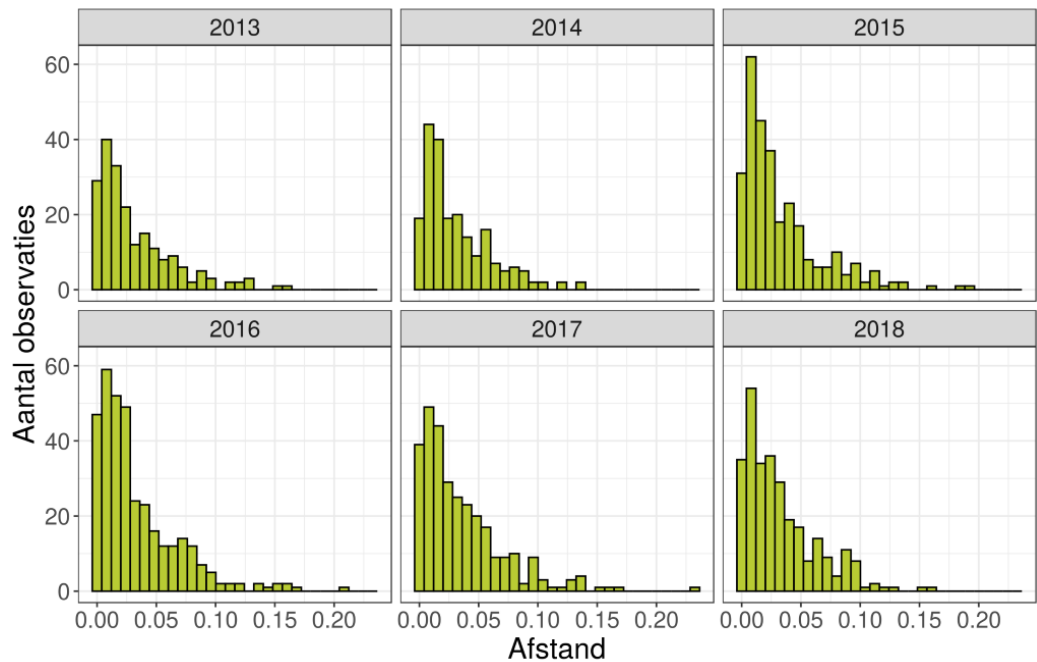
In de hoofdtekst werd gebruikgemaakt van deterministisch koppelen (zie *Paragraaf 3.3.3* voor uitleg over de verschillende technieken). Ongevallen in BRON en in de ambulanceregistratie werden gekoppeld als de gemeente en het jaar overeenstemden, en als de afstand (uitgedrukt in breedtegraden) minder dan een bepaalde eenheid was. Het aantal koppelingen kan worden vergroot door de toegestane afstand in locatie van het ongeval (BRON) en locatie van de ambulance te vergroten; dit zorgt er ook voor dat een groter aandeel van de regels die in beide bestanden overeenkomen gekoppeld gaat worden. Mogelijk is het daarom beter om gebruik te maken van technieken die uit de data schatten hoe groot de kans is dat twee regels uit verschillende bestanden bij elkaar horen, en vervolgens combinaties zo te kiezen dat elke regel in het ene bestand slechts aan één regel uit de andere bron gekoppeld wordt (probabilistisch koppelen met een-op-een-relaties; voor achtergrond, zie Christen, 2012).

Voor deze koppeling maken we hier gebruik van het fastLink-softwarepakket (Enamorado, Fifield, & Imai, 2018). Om de koppeling in een beperkte rekentijd te kunnen uitvoeren, worden vooraf die kandidaatparen in beide bestanden geselecteerd met een redelijke kans op een koppeling. In dit geval is hiervoor gebruik gemaakt van het jaar van registratie. In het proces dwingen we ook een een-op-een-koppeling af, wat binnen het pakket met het 'greedy' algoritme (snel, maar niet altijd optimaal) wordt uitgevoerd. De kwaliteit van de koppeling beoordelen we met een vergelijking van gekoppelde gemeenten en de gekoppelde afstand.

Alle letselongevallen uit BRON (N = 1.638) konden daarbij gecombineerd worden met een van de regels uit de ambulancegegevens (N = 7.002). Daarbij zorgt fastLink steeds voor een perfecte koppeling op gemeente, en staat ruimte toe bij de koppeling op de afstand in locatie binnen deze gemeente (zie *Afbeelding H.1 en H.2*). We hebben in deze data geen beschikking over informatie over de werkelijke koppelingen en daarom zal het lastig zijn de kwaliteit van de koppeling in een getal uit te drukken. De afstanden zijn relatief klein en de gemeenten stemmen steeds overeen, maar dit is geen garantie dat de gecombineerde gegevens uit de twee bronnen bij elkaar horen.



**Afbeelding H.1.** Overeenstemming in gekoppelde gemeenten tussen ambulanceregistratie (open data Flevoland) en BRON (alleen letselongevallen) op basis van fastLink.



**Afbeelding H.2.** Voorkomen van afstanden tussen de gekoppelde registraties per jaar op basis van fastLink.

## Voorbeeld van koppelen met afstandsfunctie

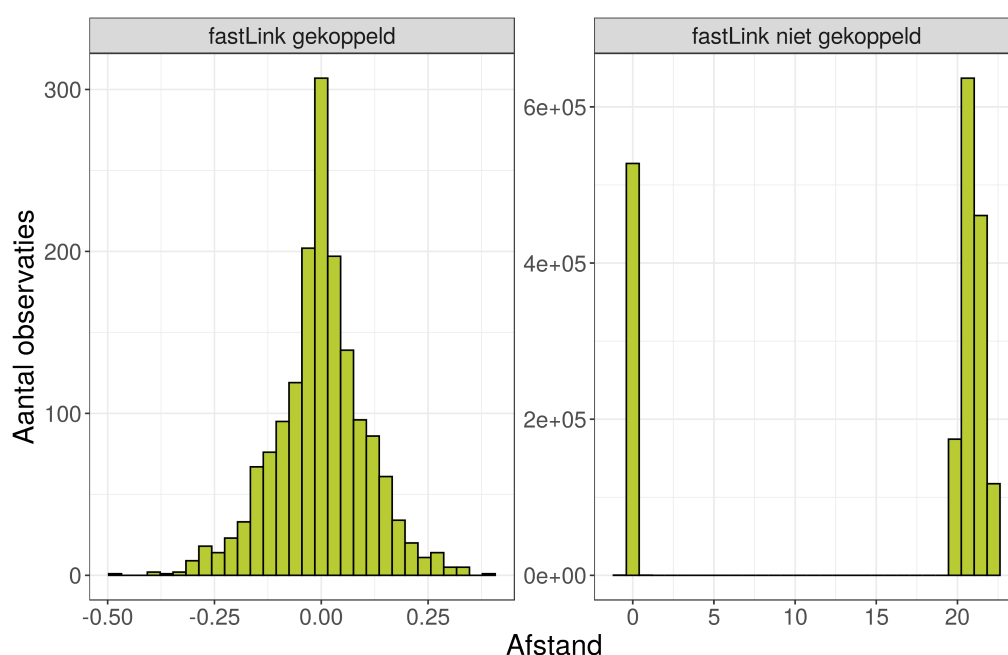
Het is ook mogelijk om te koppelen met een afstandsfunctie (zie *Paragraaf 3.3.3*). Hierbij worden gewichten gekozen om de verschillende kenmerken (overlap in gemeente, jaar, locatie) te combineren tot een afstandsmaat die aangeeft hoe waarschijnlijk een koppeling is. Een manier om deze gewichten te schatten, is door gebruik te maken van het resultaat van de probabilistische koppeling. Daartoe genereren we de set van alle kandidaatparen, en geven we voor elk aan of deze volgens de probabilistische koppeling gekoppeld worden of niet. De variabele die op deze manier wordt verkregen, kan vervolgens worden voorspeld met een logistische regressie met als voorspellers de overeenstemming van kenmerken van elk paar. De probabilistische koppeling gebruiken we hier dus als de 'gouden standaard' voor het bepalen van de optimale gewichten. Maar let op: we hebben dus geen informatie welke informatie daadwerkelijk bij elkaar hoort, dus het gaat hier vooral om een demonstratie van hoe de techniek kan werken. Met een logistische regressie bepalen we de optimale waarden van de gewichten die voor elk paar als het ware 'voorspelt' of deze probabilistisch gekoppeld wordt, dus met als voorspellers de overeenstemming van kenmerken van elk paar. Ook hier gebruiken we alleen de BRON-gegevens met ten minste één slachtoffer tezamen met de resultaten van de eerder uitgevoerde fastLink-koppeling.

Voor de logistische regressie gebruiken we het 'Caret'-pakket binnen R. Voor de 'training' van welke paren bij elkaar zouden horen, wordt eerst een dataset aangemaakt met daarin de volgende variabelen:

1. wel of niet gekoppeld met fastLink (de te voorspellen variabele);
2. een dummyvariabele die aangeeft of de gemeente wel of niet overeenkomt;
3. de afstand op de breedtegraad;
4. de afstand op de lengtegraad.

In deze dataset zitten voor elk jaar alle mogelijke combinaties van een ambulance-ID en een BRON-ID en de daarbij behorende kenmerken en de genoemde afgeleide variabele 1 t/m 4 die hierboven genoemd zijn. De dataset wordt vervolgens gesplitst op 80% van de gegevens voor de training en 20% voor de validatie. Omdat de 80% die voor de training wordt gebruikt bijzonder veel voorbeelden heeft van niet-gekoppelde ambulance en BRON-ID's, wordt een selectie gemaakt, waarbij alle voorbeelden van gekoppelde gegevens worden gebruikt en een willekeurige trekking van hetzelfde aantal voorbeelden van niet-gekoppelde gegevens. Dit wordt gedaan om te voorkomen dat de logistische regressie alles als 'niet gekoppeld' gaat voorspellen (omdat dit een zeer hoog percentage 'correct' gaat opleveren).

Voor alle voorbeelden in de oorspronkelijke dataset kan nu de afstand als een gewogen som van de bijdrage van de variabelen worden bepaald en de verdeling hiervan kan worden vergeleken voor wel door fastLink gekoppelde kandidaatparen en niet-gekoppelde kandidaatparen (zie *Afbeelding H.3*).



**Afbeelding H.3.** Afstanden volgens een geschatte afstandsfunctie op basis van overlap in gemeente, en hoogtegraad- en breedtegraadafstanden, gesplitst voor wel en niet door fastLink gekoppelde kandidaatparen in de open ambulanceregistratie van Flevoland en letselongevallen in BRON.

Het resultaat laat zien dat de probabilistische koppeling weliswaar goede koppelingen oplevert (jaar komt overeen en afstand is klein), maar ook dat een groot deel van de kandidaatparen die niet door de afstandsfunctie wordt gekoppeld (rechter deel van *Afbeelding H.3*, linker piek) ook een kleine afstand hebben. Anders gezegd: Alle 1.638 gekoppelde paren hebben een kleine afstand maar ook hebben 527.985 uit 1.389.428 kandidaatparen een kleine afstand. Als we dus alles zouden koppelen met een kleine afstand, dan verkrijgen we een aanzienlijk aantal voorstellen voor koppelingen die waarschijnlijk in werkelijkheid geen koppelingen zijn.

Alles bij elkaar, kunnen we concluderen dat alleen jaar en hoogte- en breedtegraad (en impliciet de gemeente) onvoldoende zijn om met zekerheid te kunnen zeggen welke BRON- en welke ambulancegegevens bij elkaar horen.

# Ongevallen voorkomen Letsel beperken Levens redden

## **SWOV**

**Instituut voor Wetenschappelijk Onderzoek Verkeersveiligheid**

Postbus 93113

2509 AC Den Haag

Bezuidenhoutseweg 62

070 – 317 33 33

info@swov.nl

www.swov.nl

 [@swov\\_nl](#) / [@swov](#)

 [linkedin.com/company/swov](https://www.linkedin.com/company/swov)