

ARTIKEL

Verschillen tussen buurtclusters in de kans op woningbrand en samenhangende buurtkenmerken: een verkenning*

Max Geelen & Ike Kroesbergen

Dit verkennende onderzoek beschrijft de kans op woningbrand in buurten in een regio die sterk varieert in verstedelijking. Met clusteranalyse op buurtkenmerken uit open datasets zijn buurtclusters onderscheiden die verschillen in de kans op woningbrand. In de buurtclusters met de grootste kans op woningbrand zijn in een lineair regressiemodel de meest samenhangende buurtkenmerken: percentage woningen in bezit van particuliere verhuurders, percentage bewoners met een migratieachtergrond, percentage huishoudens met een inkomen rond of onder het sociaal minimum, percentage eenpersoonshuishoudens en percentage rokers. Deze resultaten geven richting aan vervolgonderzoek. De uitkomsten geven geen oorzaak-gevolgrelaties en gelden niet voor individuele huishoudens.

1 Inleiding

Het programma Brandveilig Leven blijkt na afweging van de maatschappelijke kosten en baten effectief wanneer toegepast in buurten met een verhoogde kans op woningbrand (Akker et al., 2010). Ook andere onderzoeken (Clare et al., 2012; Eysink Smeets et al., 2015; Guldåker, 2020) onderbouwen de toepassing van brandpreventie juist in buurten met een verhoogde kans op woningbrand. Dat roept de vraag op hoe buurten met een verhoogde kans op woningbrand te onderscheiden op buurtkenmerken. Het gaat hierbij om een verkenning van buurten en hun kenmerken, en niet om een causaal onderzoek waarbij oorzaak-gevolgrelaties of brandoorzaken worden onderzocht. Bij onderscheidende kenmerken tussen buurten kan gedacht worden aan kenmerken van de woning, de woonomgeving of de buurtbewoners.

Studies naar de kans op woningbrand kunnen op verschillende manieren zijn uitgevoerd en betreffen zogenaamde ecologische studies op het niveau van een geo-

* De auteurs danken Rob van Bussel, Metha de Heer, Floor van Lintel en Dirk Suchy van de Veiligheidsregio Midden- en West-Brabant, Steven Troost van de Veiligheids- en Gezondheidsregio Gelderland-Midden en Leonard Vanbrabant van de GGD West-Brabant voor hun hulp bij het onderzoek. Wij danken de programmamanagers Louise Schneider en Jolande van Balen van de Veiligheidsregio Midden- en West-Brabant voor hun steun bij de facilitering van dit onderzoek. Naast de Veiligheidsregio Midden- en West-Brabant heeft ook de GGD West-Brabant het onderzoek gefaciliteerd.

grafisch gebied en onderzoek op het individuele niveau van woningen of huishoudens. De uitkomsten van deze studies vullen elkaar aan; deze uitkomsten kunnen niet van het gebiedsniveau vertaald worden naar het individuele niveau of omgekeerd (Mackenbach, 2000). Het eerste kenmerk dat van belang is voor de kans op woningbrand in een gebied is het aantal inwoners in het gebied of constructen daarvan zoals het aantal huishoudens (Turner et al., 2017; Van den Berg et al., 2017). Met een onderzoek in Amsterdam als inspiratiebron (Van den Berg et al., 2017) is eerst nagegaan of de buurten in de onderzoeksregio te onderscheiden zijn in de kans op woningbrand wanneer het aantal inwoners per buurt als kenmerk wordt gebruikt. Dit geeft met een visuele inspectie op een geografische kaart onvoldoende onderscheid tussen buurten in de minder verstedelijkte gebieden. De eerste uitdaging in dit onderzoek is om buurten te onderscheiden in een regio met een sterke variatie in verstedelijkte en meer landelijke buurten, met méér kenmerken dan het aantal inwoners per buurt of constructen daarvan zoals het aantal huishoudens. In zogenaamd ecologisch onderzoek van D'Elia et al. (2016) naar de kans op woningbrand in een buurt zijn als relevante buurtkenmerken gevonden die positief hiermee samenhangen: het percentage huurhuizen, het percentage onderkomens met een oudere bewoner (≥ 65 jaar) en het percentage onderkomens in een buurt met een bewoner die werkloos is. Ook de woningwaarde van de huizen in de buurt en het verhuisgedrag (het percentage bewoners dat de laatste vijf jaar is verhuisd) zijn gevonden als relevant, met een negatieve samenhang met de kans op woningbrand in de buurt. Ook het percentage bewoners met een migratieachtergrond, het percentage bewoners met een lage sociaaleconomische status en het bouwjaar van de huizen in een buurt zijn gevonden als relevante buurtkenmerken (Lehna et al., 2016), evenals het percentage eenoudergezinnen (Beaulieu et al., 2019) en het percentage rokers (Diekman et al., 2008).

De studies die op het individuele niveau van woonadres of huishouden zijn uitgevoerd, vinden in westerse landen als relevante kenmerken voor de kans op woningbrand in een woning of huishouden: man, kind onder de 5 jaar, 'niet-werkende' huishoudens, roken, laag inkomen, niet-particulier eigendom, appartement en gebouw in slechte conditie (Turner et al., 2017). Naber et al. (2019) hebben in huishoudens met een lagere sociaaleconomische status een hogere kans op woningbrand gevonden in een studie met data van verschillende regio's in Nederland. Nilson et al. (2015) vonden deze aanwijzingen ook in Zweden, maar met een andere samenhang: juist de huishoudens met een hoog opleidingsniveau bleken een hogere kans op woningbrand te hebben, evenals huishoudens met 6-12-jarige kinderen en huishoudens afkomstig van buiten de noordelijke landen. Bij onderzoek naar schade door een woningbrand (Verbond van Verzekeraars, 2021) of overlijden van bewoners door een woningbrand kunnen specifieke kenmerken naar voren komen, waaronder de oudere leeftijd en gezondheidsbeperkingen zoals in mobiliteit (Brandweeracademie, 2018, 2021). De tweede uitdaging in dit onderzoek heeft te maken met het betrekken van databronnen over leefstijl en gezondheid. Leefstijl zoals roken en gezondheid(sbeperkingen, zoals in mobiliteit) lijken relevant voor de kans op woningbrand (o.a. Diekman et al., 2008; NIPV, 2022), maar zijn als kenmerken van menselijk gedrag nog nauwelijks betrokken in onderzoek (Lehna et al., 2016; Beaulieu et al. 2019; Naber et al., 2019; School et al., 2021). In dit verken-

nende onderzoek, waarin de samenhang tussen buurtkenmerken en de kans op woningbrand in buurten wordt onderzocht, worden ook leefstijl- en gezondheidskenmerken over de buurtbewoners als buurtkenmerk betrokken.

Omdat voor brandpreventieve programma's aanbevolen wordt om deze toe te passen in gebieden met een hoge kans op woningbrand, is de probleemstelling in dit onderzoek of buurttypen te onderscheiden zijn met kenmerken die verder gaan dan het aantal inwoners en die relevant zijn voor de kans op woningbrand in het buurttype. De onderzoeksvragen zijn:

- Kunnen buurten in een regio die sterk wisselt in de mate van verstedelijking geclusterd worden tot een aantal buurttypen op basis van buurtkenmerken die verder gaan dan het aantal inwoners in de buurt en verschillen deze buurttypen in de kans op woningbrand?
- Welke buurtkenmerken hangen in ieder van deze buurttypen het meest samen met de kans op woningbrand?

2 Methode

In dit onderzoek wordt lineaire regressie toegepast als verkenning van de samenhang tussen de kans op woningbrand in een buurt en buurtkenmerken. Hierbij zijn we voornamelijk geïnteresseerd in de buurtkenmerken (attributen) die het regressiemodel als meest betekenisvol beschouwt. We passen een aantal varianten van regressiemodellen toe, die beoordeeld worden op hoe goed ze in staat zijn om de kans op woningbrand in een buurt te schatten. Het resultaat is een getraind regressiemodel waarmee we kunnen bepalen welke buurtkenmerken het meest samenhangen met de kans op woningbrand in de buurt.

In dit onderzoek volgen we het procesmodel van *Cross Industry Standard for Data Mining* (CRISP-DM, 2022). De implementatie (Deployment)-fase is nog niet aan de orde en laten we hier buiten beschouwing. We beschrijven de verschillende stappen van dit model.

2.1 Context van data

De eerste stap is om een zo helder mogelijk begrip van de data te ontwikkelen. Hiervoor is het essentieel om de context van de data goed te begrijpen. In de Gemeenschappelijke Meldkamer van politie, brandweer en ambulance worden meldingen aangenomen door de centralisten. Een willekeurig persoon kan via het alarmnummer de centrale meldkamer bereiken en een melding maken van een incident. Samen met de centralist wordt er beoordeeld of er één of meer hulpdiensten opgeroepen worden om ter plaatse te gaan. In dit onderzoek richten we ons op de incidenten waar de brandweer is opgeroepen om uit te rukken. Een centralist classificeert een incident in het Geïntegreerd Meldkamer Systeem (GMS). Deze classificatie wordt in dit onderzoek gebruikt om te bepalen of een incident een brand in een woning of een gebouw met een woonfunctie betreft. Gedurende het incident blijft de centralist de inzet monitoren. Zodra de hulpdiensten klaar zijn met de inzet, wordt het betreffende incident afgesloten door de centralist. Een

eindrapportage wordt opgesteld; dit gebeurt niet bij bijvoorbeeld onterechte meldingen of testmeldingen.

2.2 Datadefinities

De volgende stap is om de datadefinities vast te leggen. *Kans op woningbrand* is gedefinieerd als het aantal woningbranden per 1000 huishoudens in een buurt gedurende 9 jaar (2013 t/m 2021). Een *woningbrand* is gedefinieerd als een melding bij de Gemeenschappelijke Meldkamer vastgelegd in het GMS van de brandweer, met een inzet van in ieder geval de brandweer. Bij deze woningbranden kan letsel of schade zijn ontstaan; in de dataset zijn geen gegevens bekend over eventueel letsel en schade aan eigendommen. Het betreft unieke incidenten (dubbele meldingen zijn samengevoegd), waarvan de status is afgesloten en bovendien is er een eindrapportage gemaakt. In dit onderzoek beoordelen we een woningbrand als: 1) incident geclassificeerd als woningbrand of 2) gebouwbrand in een gebouw met woonfunctie. Het gebruikte databestand is geschoond voor oefen-, test- of onterechte meldingen en/of incidenten waarbij de brandweer niet is uitgerukt. Alle incidenten die aan de definitie voldoen vanaf 1 januari 2013 tot en met 31 december 2021 zijn meegenomen in de analyses. In totaal gaat het in de onderzoeksregio in de genoemde periode om 3749 woningbranden. De incidenten zijn geaggregeerd op CBS-buurtniveau. De gebruikte *buurtindeling* is afkomstig van het CBS (2016), in totaal gaat het om 679 buurten. Verder gebruiken we de *buurtkenmerken* (2016) uit een openbare dataset van het CBS met kerncijfers over wijken en buurten. Buurten met minder dan 100 huishoudens, bijvoorbeeld industrieterreinen waar weinig huishoudens zijn of landelijke gebieden waar nauwelijks bebouwing is, zijn niet meegenomen, omdat er voor deze gebieden meerdere gegevens ontbreken in deze dataset van het CBS. Om deze reden zijn er 194 buurten afgevallen en zijn 485 buurten betrokken in de analyses. Kenmerken over leefstijl en gezondheid van de buurtbewoners komen uit de Gezondheidsmonitor 2016 (GGD'en, RIVM & CBS, 2016). Dit betreft een vierjaarlijkse enquête die is toegestuurd aan een steekproef van inwoners van 19 jaar en ouder. De uitkomsten hiervan zijn geschat op CBS-buurtniveau en zijn openbaar beschikbaar (Van de Kassteele et al., 2017).

2.3 Datavoorbereiding

De buurtkenmerken aanwezig in de open data van het CBS (2016) en GGD'en, RIVM en CBS (2016) vormen de uitgangssituatie voor dit onderzoek. Resultaat hiervan is een dataset met daarin de buurten met meer dan 100 huishoudens en een aantal kwantitatieve buurtkenmerken (attributen). In dit onderzoek ligt de focus op buurtkenmerken die verder gaan dan constructen van het aantal inwoners. Attributen die direct of indirect het aantal inwoners in de buurt representeren, zijn niet meegenomen in de analyse. Daarom zijn attributen als de afstand tot bepaalde voorzieningen, het percentage adressen aangesloten op stadsverwarming en ook de oppervlakte van water en land verwijderd. Bijlage II geeft een overzicht van alle attributen die betrokken zijn in dit onderzoek. Ten slotte zijn de attributen uitgedrukt in percentages of buurtgemiddelden; dit resulteert in een dataset met enkel continue attributen. In sommige gevallen ontbreekt er een aantal waardes in de dataset. Attributen waar meer dan 5% van de waardes missen, zijn uit de data-

set verwijderd. De overige missende waardes worden door middel van data-imputatie ingevuld. Hiervoor is het multivariate imputation by chained equations (MICE)-pakket beschikbaar voor R gebruikt en specifiek de PMM (Predictive Mean Matching)-methode. Voor het beste resultaat van de data-imputatie zijn het aantal inwoners en het aantal huishoudens in een buurt pas na voltooiën van de data-imputatie verwijderd uit de dataset. Resultaat is een dataset waar geen missende waardes zijn. Vervolgens is de afstand (of gelijkenis) bepaald van elke buurt tot elkaar door middel van de afstandsmatrix. Als eerste zijn de attributen genormaliseerd, dit om attributen gemeten in verschillende eenheden of schalen met elkaar te kunnen vergelijken. De afstandsmatrix is vervolgens berekend aan de hand van de euclidische afstand. Andere transformatiestappen zijn niet nodig, omdat de dataset geen andere datatypen bevat dan continue attributen.

2.4 Modelleren

Als uitgangspunt om het aantal woningbranden in een buurt te schatten is een spreidingsdiagram gemaakt tussen het aantal woningbranden in een buurt en het aantal huishoudens in een buurt. Daarbij passen we lineaire regressie toe om het lineaire verband te kwantificeren (nul- of referentiemodel). Als maatstaf voor de kwaliteit van het model gebruiken we de RMSE (wortel van de gemiddelde kwadratische fout). We analyseren ook de MAE (gemiddelde absolute fout) in de schatting van woningbrand in een buurt. Vervolgens zijn we benieuwd of we buurttypen kunnen onderscheiden zonder het aantal huishoudens in de buurt mee te nemen. We passen een k-Means clustering van buurten toe. k-Means clustering (MacQueen, 1967) is een clusteranalyse-algoritme dat objecten groepeerd in k aparte groepen aan de hand van de attribuutwaardes. k is een positief geheel getal dat het aantal clusters aangeeft en moet vooraf worden opgegeven. We bepalen de waarde k aan de hand van de scree-test; $k = 4$ is overwogen, $k = 5$ lijkt het meest geschikt. In het vervolg van de analyses zijn deze 5 buurtclusters toegepast. Buurten ingedeeld in dezelfde groep hebben vergelijkbare attribuutwaardes. Willekeurige buurten binnen een cluster zullen meer op elkaar lijken dan willekeurige buurten vanuit verschillende clusters. De afstandsmatrix, zoals beschreven in paragraaf 2.3, is input voor het k-Means cluster algoritme. We beoordelen of de buurtclusters statistisch van elkaar verschillen aan de hand van een variantieanalyse op basis van het aantal woningbranden per 1000 huishoudens in 9 jaar (2013 t/m 2021). In de nulhypothese stellen we dat er geen verschil is tussen het aantal woningbranden per 1000 huishoudens in 9 jaar in de clusters. We drukken de ratio van buurt n als volgt uit:

$$r_n = \frac{\#woningbranden_n}{\left(\frac{\#huishoudens_n}{1000}\right)}$$

De gemiddelde ratio binnen cluster k drukken we uit als: \bar{r}_k

We formuleren de hypothese als volgt:

$$H_0 : r_1 = r_2 = r_3 = r_4 = r_5$$

$$H_1 : \text{ratio's niet allemaal gelijk.}$$

De verschillen tussen de clusters maken we visueel met een spreidingsdiagram per buurtcluster. Hierin staat voor ieder buurtcluster het aantal woningbranden tegenover het aantal huishoudens, in combinatie met de lineaire regressie. Aan de hand van de richtingscoëfficiënt en de bijbehorende lineaire lijn is vastgesteld of er relatief meer of minder woningbranden plaatsvinden in de betreffende buurtclusters. Zo ontstaat een lineair regressiemodel per buurtcluster (model 1), dat wordt vergeleken met het nulmodel op basis van de RMSE. We stellen dat model 1 beter presteert dan model 0 als de RMSE lager is.

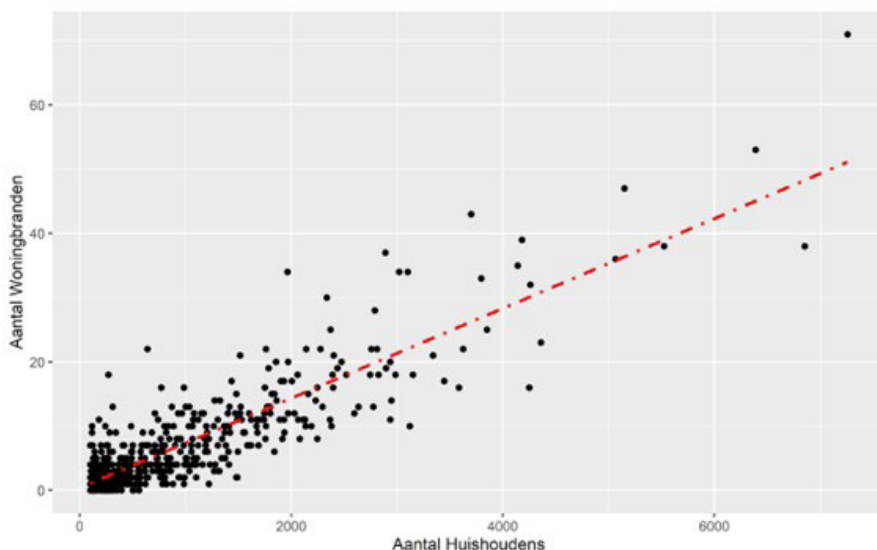
Vervolgens zijn we geïnteresseerd in welke attributen het meest samenhangen met de kans op woningbrand in het betreffende buurtcluster. Hiervoor trainen we een Lasso-regressiemodel dat gebruikmaakt van regularisatie. Hierdoor wordt het model gestraft wanneer het een extra coëfficiënt gebruikt voor de schatting. Hiermee voorkomt het model ‘overfitting’ door de flexibiliteit te beperken. Voordeel van het Lasso-regressiemodel is dat het in staat is om attributen volledig te elimineren. Hierdoor blijft slechts een beperkt aantal attributen over, waardoor het regressiemodel gemakkelijker te interpreteren is. We laten het aantal inwoners of huishoudens (of een construct daarvan) weg uit het regressiemodel, omdat we juist op zoek zijn naar de andere attributen die samenhangen met de kans op woningbrand in een bepaald buurttype. Aan de hand van het getrainde model kunnen we de richting (positieve of negatieve samenhang) en het belang (de waarde van de regressiecoëfficiënt) van de attributen bepalen voor de schatting van woningbrand in een bepaald buurttype. De afhankelijke variabele, het aantal woningbranden in een periode van 9 jaar, is rechts scheef verdeeld. Dit wil zeggen dat er relatief veel buurten zijn met weinig woningbranden en relatief minder buurten met veel woningbranden. Afhankelijk van de resultaten van het Lasso-regressiemodel zijn mogelijk nog transformatiestappen nodig om te voldoen aan de aannames van lineaire regressie. Er is hyperparameter tuning toegepast in combinatie met een 10-fold cross-validatie om het Lasso-model te trainen voor elk buurtcluster. Het model met de beste hyperparameter is gekozen aan de hand van de laagste RMSE. Voor de hyperparameter tuning is een tuning raster (grid) gebruikt met een lambda 0,0001 tot 1 in 100 stappen. Het resultaat is een Lasso-regressiemodel per buurtcluster (model 2) met de meest samenhangende buurtkenmerken voor de kans op woningbrand in dit buurtcluster.

3 Resultaten

3.1 Lineaire regressie woningbranden versus aantal huishoudens (model 0)

In figuur 1 is het nul- of referentiemodel (spreidingsdiagram) weergegeven met hierin het aantal woningbranden in 9 jaar (2013 t/m 2021) per buurt versus het aantal huishoudens (2016) in de betreffende buurt. De lineaire lijn in de figuur geeft de lineaire regressie weer waarbij het aantal huishoudens in een buurt een verklaring is voor het aantal woningbranden in een buurt. Buurten aan de bovenkant van deze lijn hebben gemiddeld gezien, in vergelijking met het aantal huishoudens in een buurt, meer woningbranden. Het tegenovergestelde geldt ook: buurten aan de onderkant van de lineaire lijn hebben gemiddeld gezien minder woningbranden. Met name buurten met een grote verticale afstand tot de lineaire regressielijn zijn interessant om verder te onderzoeken.

Figuur 1 Spreidingsdiagram van buurten – aantal woningbranden 2013 t/m 2021 t.o.v. aantal huishoudens 2016

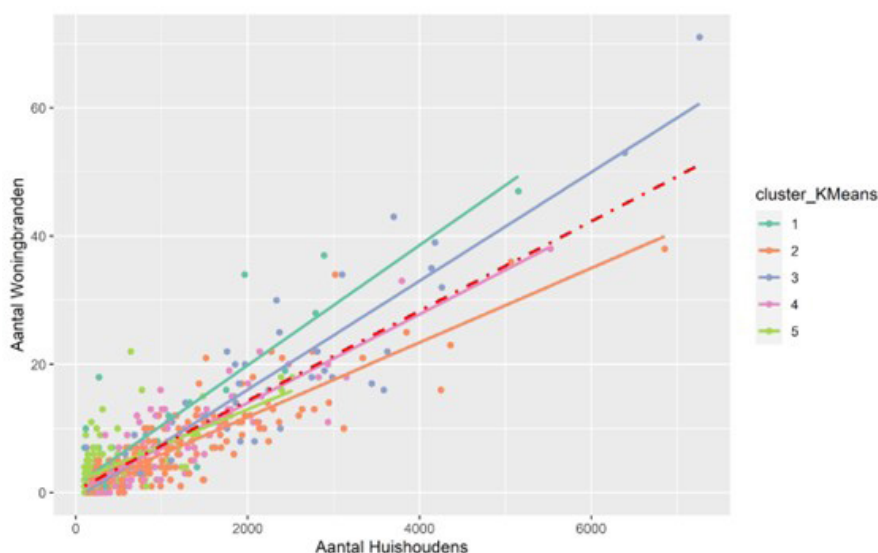


Het nulmodel heeft een R^2 van 0,75, dat wil zeggen dat 75% van de variantie in het aantal woningbranden in buurten verklaard kan worden aan de hand van het aantal huishoudens in de buurt. Dit is een sterk verband. We beoordelen hoe goed het nulmodel in staat is het aantal woningbranden te schatten aan de hand van RMSE (wortel van de gemiddelde kwadratische fout), deze bedraagt 4,12. De MAE (gemiddelde absolute fout) bedraagt 2,87. Dat wil zeggen dat we gemiddeld gezien over 9 jaar 2,87 meer of minder woningbranden schatten aan de hand van het aantal huishoudens in een buurt dan daadwerkelijk hebben plaatsgevonden.

3.2 Lineaire regressie woningbranden versus aantal huishoudens per buurtcluster (model 1)

In figuur 2 zijn de buurten en het aantal woningbranden opnieuw weergegeven in een spreidingsdiagram, maar dan per buurtcluster. Buurten die zijn ingedeeld in hetzelfde cluster worden aangeduid met eenzelfde kleur. Voor elk cluster is het bijbehorende lineaire regressiemodel (model 1) weergegeven in de bijbehorende kleur. Het lineaire nulmodel is weergegeven met de rode stippellijn.

Figuur 2 Spreidingsdiagram van buurten voor ieder buurtcluster – aantal woningbranden 2013 t/m 2021 t.o.v. aantal huishoudens 2016



De regressiecoëfficiënt (hoek) van het lineaire model geeft de verhouding aan tussen het aantal woningbranden en het aantal huishoudens. Hoe hoger de coëfficiënt, hoe steiler de lijn in de figuur. De coëfficiënt geeft aan hoe sterk het aantal woningbranden stijgt met een toename van het aantal huishoudens. De buurtclusters verschillen van elkaar op basis van het aantal woningbranden in 9 jaar per 1000 huishoudens (ANOVA $F(4, 480) = 16,9; p < 0,001$). In tabel 1 staat de samenvattende statistiek voor model 1. Voor elk buurtcluster staan daarin het aantal buurten (n_i) in het cluster, het gemiddelde aantal woningbranden in 9 jaar per 1000 huishoudens (\bar{X}_i), de regressiecoëfficiënt (β_i) van het lineaire regressiemodel, de RMSE (wortel van de gemiddelde kwadratische fout) en de MAE (gemiddelde absolute fout). Gemiddeld is de RMSE 3,67. De schatting van de kans op woningbrand blijkt 0,45 verbeterd ten opzichte van het nulmodel, als de buurtclustering in de schatting is meegenomen. De MAE (gemiddelde absolute fout) is 2,63; deze blijkt 0,24 verbeterd ten opzichte van het nulmodel. Buurtcluster 1 en buurtcluster 3 hebben gemiddeld over 9 jaar per 1000 huishoudens meer woningbranden dan de andere buurtclusters.

Tabel 1 Resultaten lineaire regressie per buurtcluster (model 1)

i	n_i	\underline{X}_i	β_i	$RMSE_i$	MAE_i
1	32	10,16	9,35	5,11	7,85
2	171	6,85	5,80	3,33	4,04
3	58	13,97	8,48	5,09	9,92
4	93	7,87	6,90	3,37	4,16
5	131	3,88	5,62	3,08	1,62
Tot./Gem.	485	7,31		3,67	2,63

3.3 Lasso-regressie naar meest samenhangende attributen per buurtcluster (model 2)

De resultaten van de Lasso-regressie om de meest samenhangende buurtkenmerken voor woningbrand per buurtcluster te bepalen, staan in tabel 2. Voor elk buurtcluster zijn de gemiddelde $RMSE$, R^2 en de MAE weergegeven voor de 10 folds van het beste model. Het beste model is bepaald aan de hand van de waarde van parameter Lambda (λ_i) op basis van $RMSE$, waarbij een lagere waarde van $RMSE$ gewenst is. Uit een visuele inspectie concluderen we dat de residuen normaal verdeeld lijken te zijn, waardoor verdere transformatiestappen niet nodig blijken. Merk op dat het Lasso-model een hoge R^2 heeft voor cluster 1, namelijk 0,71 (zie tabel 2). Cluster 1 is het buurtcluster met een bovengemiddeld aantal woningbranden. De R^2 geeft aan hoeveel van de variantie in woningbranden in het regressie-model door de buurtkenmerken verklaard wordt. Voor cluster 3, ook een buurtcluster met een bovengemiddeld aantal woningbranden, is de R^2 0,36.

Tabel 2 Resultaten Lasso-regressie per buurtcluster (model 2)

i	λ_i	R_i^2	$RMSE_i$	MAE_i
1	0,63	0,71	7,82	6,40
2	0,86	0,11	6,21	4,60
3	1	0,36	11,65	8,98
4	1	0,24	6,53	5,08
5	0,37	0,20	3,65	2,62
Gem.			6,10	4,11

Hieronder bespreken we de resultaten van de getrainde modellen. Hierbij geven we aan welke buurtkenmerken het meest samenhangen met de kans op woningbrand in het betreffende buurtcluster. Deze analyses gaan over het buurtniveau en niet over branden op een woonadres of bij een individueel huishouden en het gaat niet om causale verbanden of brandoorzaken. De top 5 meest samenhangende attributen per buurtcluster zijn opgenomen in tabel 4 met de bijbehorende regressiecoëfficiënten. In tabel 3 is een korte uitleg gegeven van de attributen uit tabel 4. Bijlage I geeft het totaaloverzicht van de attributen die een bijdrage leveren aan de schatting van de kans op woningbrand per buurtcluster. We lichten de top 5 attributen toe uit buurtcluster 1 en 3, omdat in deze buurtclusters de kans op woningbrand het hoogst is. Met name buurtcluster 1 is interessant, omdat het regressiemodel in dit buurtcluster het best functioneert. Voor buurtcluster 1 is het kenmerk met de hoogste coëfficiënt: het percentage woningen in een buurt in het bezit van verhuurders anders dan woningcorporaties (denk aan particuliere verhuurders). Buurten met een hogere kans op woningbrand hebben doorgaans ook een hoger percentage woningen in bezit van verhuurders anders dan woningcorporaties. De positieve relaties hebben in tabel 4 een *. Ook zijn er drie attributen (p_{antaru} , p_{turtur} , p_{ovnw}) die betrekking hebben op het percentage inwoners met een (niet-westerse) migratieachtergrond. Buurten met een hogere kans op woningbrand hebben doorgaans ook een hoger percentage inwoners met een betreffende migratieachtergrond. Het percentage bedrijven uit de zakelijke dienstverlening heeft een negatieve relatie met het aantal woningbranden in een buurt (de negatieve relaties worden in tabel 4 aangegeven met **). Buurten waar meer woningbranden plaatsvinden, hebben doorgaans een lager percentage KvK-inschrijvingen uit de zakelijke dienstverlening. In buurtcluster 3 heeft naast twee attributen over migratieachtergrond ook het percentage huishoudens met een inkomen onder of rond het sociaal minimum een positief verband met woningbranden in een buurt. Dit geldt ook voor het percentage eenpersoonshuishoudens en het percentage rokers in een buurt.

Tabel 3 *Korte beschrijving van attributen die terugkomen in tabel 4*

p_ov_hw	Percentage woningen in bezit van overige verhuurders (dus geen woningcorporaties)	p_hh_osm	Percentage huishoudens met inkomen onder of rond sociaal minimum
p_antaru	Percentage inwoners met een migratieachtergrond uit Nederlandse Antillen en Aruba	p_l_p_hh	Percentage eenpersoonshuishoudens
p_tur	Percentage inwoners met een migratieachtergrond uit Turkije	roker_10	Percentage inwoners dat rookt
p_ov_nw	Percentage inwoners met een migratieachtergrond uit overige niet-westerse landen	p_lgezw	Percentage eengezinswoningen
p_bed_mn	Percentage bedrijfsvestigingen met SBI: Zakelijke dienstverlening	p_gehuwd	Percentage inwoners dat is getrouwd
p_45_64	Percentage inwoners tussen de 45 en 65 jaar oud	HoogRisicoOpAngstOfDepressie_20	Percentage inwoners met een hoog risico op een angststoornis of depressie
p_hh_z_k	Percentage huishoudens zonder kinderen	eenzaam_21	Percentage inwoners dat zich ernstig eenzaam voelt
p_e_o_w	Percentage woningen waarvan eigendom onbekend is	p_suri	Percentage inwoners met een migratieachtergrond uit Suriname
ErnstigBeperktVanweGezondheid_13	Percentage inwoners dat zegt ernstig te worden beperkt door hun gezondheid. (minstens 1 OESO-beperking)	Overgewicht_8	Percentage inwoners met een BMI $\geq 25,0$ kg/m ²
Mobiliteitsbeperking_17	Percentage inwoners dat zegt beperkt te zijn in hun mobiliteit vanwege gezondheid	p_w_all	Percentage inwoners met een migratieachtergrond uit een westers land
OvermatigeDrinker_6	Percentage inwoners dat overmatig drinkt (d.w.z. meer dan 14 glazen per week voor vrouwen en meer dan 21 glazen per week voor mannen)		

Tabel 4 Top 5 attributen in combinatie met coëfficiënt per buurtcluster (model 2)

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
p_ov_hw	p_45_64	p_antaru	p_gezw	p_hh_z_k
5,01*	-0,76**	3,81*	-1,03**	-0,53**
p_antaru	p_hh_z_k	p_hh_osm	p_gehuwd	Overge- wicht_8
4,90*	-0,63**	2,93*	-0,98**	-0,53**
p_tur	p_e_o_w	p_lp_hh	HoogRisi- coOpAng- stOfDepres- sie_20	p_w_all
3,20*	0,36*	2,20*	0,17*	0,53*
p_ov_nw	ErnstigBe- perktVanwe- Gegezond- heid_13	roker_10	eenzaam_21	Overmatige- Drinker_6
0,81*	0,28*	1,59*	0,08*	0,37*
p_bed_mn	Mobiliteitsbe- perking_17	p_tur	p_suri	p_hh_osm
-0,69**	0,24*	0,90*	0,06*	0,35*

* Er is een positieve relatie. ** Er is een negatieve relatie.

4 Discussie en conclusie

Dit onderzoek past een clustermethode toe aan de hand van buurtkenmerken en richt zich vervolgens op de vraag of de kans op woningbrand verschillend is in de buurtclusters. De meest samenhangende buurtkenmerken die verder gaan dan constructen van het aantal inwoners zijn voor ieder van de buurtclusters onderzocht.

De buurten in de onderzochte regio blijken geclusterd te kunnen worden in vijf buurttypen die onderling verschillen in de kans op woningbrand. De buurtclusters blijken onderling te verschillen in de buurtkenmerken die het meest samenhangen met de kans op woningbrand. In de buurtclusters met de hoogste kans op woningbrand zijn als meest samenhangende buurtkenmerken gevonden met een positieve richting van de samenhang: percentage woningen in bezit van particuliere verhuurders, percentage buurtbewoners met een migratieachtergrond, percentage huishoudens met een inkomen rond of onder het sociaal minimum, percentage eenpersoonshuishoudens en percentage rokers in de buurt.

Onderzoek op het buurtniveau tussen de kans op woningbrand en de buurtkenmerken is geschikt voor vraagstellingen zoals in dit onderzoek naar onderscheid tussen buurten. De resultaten van de regressiemodellen uit dit onderzoek geven geen oorzaak-gevolgrelaties en zijn niet geldig voor het individuele niveau van een woonadres of huishouden. Het omgekeerde is ook het geval: de samenhang die gevonden is in onderzoek waarbij de kans op brand op het individuele niveau van een adres of huishouden is gerelateerd aan kenmerken van de woning, woonomgeving en bewoners, gelden niet voor het buurtniveau en voor vraagstellingen die gericht zijn op het onderscheid tussen buurten (Mackenbach, 2000). Dat maakt dat onderzoek nodig is zowel op het ecologische niveau van de buurt als op het individuele niveau van woningen en huishoudens. De buurtkenmerken die in dit verkennende onderzoek zijn gevonden als meest samenhangend zijn eveneens in ander onderzoek gevonden dat op gebiedsniveau is uitgevoerd (Diekman et al., 2008; D'Elia et al., 2016; Lehna et al., 2016). In dit onderzoek komt 'verhuur' nog specifiek naar voren dan in ander onderzoek, omdat in buurtcluster 1 de particuliere verhuur het meest samenhangende buurtkenmerk is voor de kans op woningbrand in dit buurtcluster. Zorgvuldig communiceren over het ecologische buurtniveau van het onderzoek is van belang om misinterpretatie van de onderzoeksresultaten te voorkomen. Het gevaar van stereotypering van buurtclusters ligt op de loer.

Onder woningbrand wordt in dit onderzoek een brand in een woning of woongebouw verstaan waarbij de Gemeenschappelijke Meldkamer is gealarmeerd en de brandweer is ingezet. Hoewel er inspanningen zijn verricht om de data zo correct en compleet mogelijk te krijgen, zijn registratiefouten of incompleetheid van de data over de brandmeldingen niet geheel uit te sluiten. Zo kan de informatie die bij de melding wordt doorgegeven anders zijn dan de daadwerkelijke omstandigheden. Er is een controle op de data uitgevoerd op de woonfunctie van de locatie, waardoor de betrouwbaarheid van de data op dit punt verbeterd is. Ook betekent de benutting van data van de Gemeenschappelijke Meldkamer met een definitie van woningbrand waarbij de brandweer is ingezet, dat kenmerken van de alarmering in de definitie zijn opgenomen. Bij het wel of niet (accuraat) alarmeren kun-

nen mogelijk sociaal-demografische kenmerken een rol spelen (Tannous & Agho, 2017). Ook de omvang en ernst van een brand kunnen een rol spelen in de selectie van woningbranden die onderzocht zijn: als de bewoner zelf een beginnende brand kan blussen, wordt de brandweer mogelijk niet gealarmeerd en ingezet. Letsel en schade door de brand zijn niet in de gebruikte dataset geregistreerd, zodat dit onderzoek beperkt is tot uitspraken over de kans op woningbrand in een buurt, waarbij de brandweer is ingezet. Daar kan schade, letsel en mogelijk ook overlijden het gevolg van zijn geweest. We kunnen echter geen uitspraken doen over de kans op fatale branden of over de kans op woningbrand met letsel of schade. Ook zijn buurten waar minder dan honderd huishoudens wonen, niet meegenomen in de studie. Deze gebieden zijn mogelijk minder relevant voor onze onderzoeksvraag, denk aan industrieterreinen of buitengebieden met weinig bebouwing.

Voor de buurtkenmerken is in dit onderzoek gebruikgemaakt van bestaande open data op buurtniveau van het CBS en uit gezondheidsenquêtes van GGD'en, RIVM en CBS en de kenmerken die daarin zijn vertegenwoordigd. Op het moment van de analyses waren de datasets van 2020 nog niet beschikbaar en zijn datasets van 2016 benut. Om het aantal woningbranden zo groot mogelijk te laten zijn in de analyses is een periode van 9 jaar data over woningbranden in dit onderzoek betrokken (2013 t/m 2021). De periode voor 2013 kon niet betrokken worden, vanwege veranderingen in de codering. De buurtkenmerken zijn afkomstig uit de dataset van 2016, dit is ongeveer halverwege de periode 2013 tot en met 2021. Dit kan voor buurten die snel veranderen in de kenmerken van woningen en bewoners, bijvoorbeeld renovatie- of nieuwbouwbuurten mogelijk een vertekend beeld geven in de analyses. Vervolgonderzoek kan zich richten op herhaling van de analyses met de buurtindeling en buurtkenmerken van 2020, die recent beschikbaar zijn gekomen. Ook kan nagegaan worden of de analyses toegepast kunnen worden op grotere sets branddata waarin meer regio's vertegenwoordigd zijn. Voor de grotere regio kan dan opnieuw worden afgewogen welke attributen in het onderzoek te betrekken. In de onderzochte regio is stadsverwarming een representant van verstedelijking en om die reden weggelaten in de lijst met buurtkenmerken. Dit moet opnieuw worden beoordeeld bij onderzoek in een andere, grotere regio.

De resultaten in dit onderzoek geven aangrijpingspunten voor verder onderzoek, dat gericht kan plaatsvinden. Om de resultaten te benutten voor keuzes in buurten waar doelgroepgerichte interventies worden toegepast, is vervolgonderzoek nodig. In dit vervolgonderzoek kunnen branddata verrijkt worden met CBS-microdata of data over brandoorzaken. Vervolgens kan op het individuele (huishoudens- of woning)niveau worden geanalyseerd om een gerichte vraagstelling te beantwoorden. Voor buurtcluster 1 is uit dit onderzoek gebleken dat de kans op het ontstaan van woningbrand hoger is dan in andere buurtclusters. Een gerichte onderzoeksvraag kan bijvoorbeeld zijn of in dit buurtcluster in de woningen die particulier worden verhuurd, vaker dan gemiddeld brand is ontstaan, welke brandoorzaken er bij deze branden waren en welke kenmerken de bewoners van deze woningen hebben.

Literatuur

- Akker I., L.A.W. Tieben, J.G.H. Bos & M.J. van der Veen (2010) *Investeren in brandveiligheid: Maatschappelijke kosten-batenanalyse woningbrand*. Amsterdam: SEO Economisch onderzoek.
- Beaulieu E., J. Smith, A. Zheng & I. Pike (2019) Association between neighbourhood socioeconomic features and residential fire incidence, related casualties and children: a cross-sectional population-based study in 4 Canadian provinces. *CMAJ Open*, 7(3), 562-567, doi:10.9778/cmajo.20190079.
- Berg P.L., van den, G.A.G. Legemaate & R.D. van der Mei (2017) Increasing the responsiveness of Firefighter Services by Relocating Base Stations in Amsterdam. *Interfaces*, 47(4), 352-361, doi.org/10.1287/inte.2017.0897.
- Brandweeracademie (2018) *10 jaar fatale woningbranden onderzocht*. Arnhem: Instituut Fysieke Veiligheid.
- Brandweeracademie (2021) *Jaaroverzicht fatale woningbranden 2020*. Arnhem: Instituut Fysieke Veiligheid.
- CBS (2016) www.cbs.nl/nl-nl/maatwerk/2016/30/kerncijfers-wijken-en-buurten-2016.
- Clare J., L. Garis, D. Plecas & C. Jennings (2012) Reduced frequency and severity of residential fires following delivery of fire prevention education by on-duty fire fighters: Cluster randomized controlled study. *Journal of Safety Research*, 43, 123-28.
- CRISP_DM (2022) www.datascience-pm.com/crisp-dm-2/, geraadpleegd 28-4-2022.
- D'Elia, A.D., C.P. Muir & S.V. Newstead (2016) Factors affecting the impact and incidence of residential fires. MUARC.
- Diekman S.T., M.F. Ballesteros, L.R. Berger, R.S. Caraballo & S.R. Kegler (2008) Ecological level analysis of the relationship between smoking and residential-fire mortality. *Injury Prevention*, 14(4), 228-31, doi:10.1136/ip.2007.017004.
- Eysink Smeets M., M. Heijman & R. Postma (2015) *Gedragsverandering voor brandveilig leven: Inzichten en aanbevelingen*. Arnhem: Wetenschappelijke Raad Brandweer Nederland.
- GGD'en, CBS & RIVM (2016) *Gezondheidsmonitor Volwassenen en Ouderen, bewerking o.b.v. SMAP-methode*, <https://data.overheid.nl/dataset/5209-gezondheid-per-wijk-en-buurt-2016#description>.
- Guldåker, N. (2020) Geovisualization and Geographical Analysis for Fire Prevention. *International Journal of Geo-Information*, 9, 355, doi:10.3390/ijgi9060355.
- Kasstele J., van de, L. Zwakhals, O. Breugelmans, C. Ameling & C. Van den Brink (2017) Estimating the prevalence of 26 health-related indicators at neighbourhood level in the Netherlands using structured additive regression. *Journal of Health Geographics*, 16(1), 23.
- Lehna C., A. Speller, C. Hanchette, E. Fahey & M.B. Coty (2016) Development of a Fire Risk Model to Identify Areas of Increased Potential for Fire Occurrences *Journal of Burn Care and Research*, 37(1), 12-19, doi: 10.1097/BCR.000000000000297.
- Mackenbach, J.P. (2000) Dwalingen in de methodologie XXVI. De ecologische valkuil en zijn minder bekende tegenhanger, de atomistische valkuil. *Nederlands Tijdschrift voor Geneeskunde*, 144, 2097-100.
- MacQueen, J. (1967) Some Methods for Classification and Analysis of Multivariate Observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (p. 281-297). Berkeley: University of California Press. <http://projecteuclid.org:443/euclid.bsm/1200512992>.
- Naber N., V. Oskam, R. Piek & S. de Ruiter (2019) *Sociaaleconomisch brandrisico: Een onderzoek met CBS microdata*. Rotterdam-Rijnmond: Veiligheidsregio.

- Nilson F., C. Bonander & A. Jonsson (2015) Differences in Determinants Amongst Individuals Reporting Residential Fires in Sweden: Reports form a Cross-Sectional Study. *Fire Technology*, 51(3), doi.org/10.1007/s10694-015-0459-0.
- NIPV (2022) *Dashboard fatale woningbranden*. <https://kerncijfers.nipv.nl/mosaic/kerncijfers-veiligheidsregio-s/fatale-woningbranden>.
- R Core Team (2020) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. www.r-project.org.
- School, M., M. de Graaf, M.C. van Lieshout, E. Sanders & R. de Wit (2021) Van tellen naar voorspellen: sturen op risico's met een voorspellend wiskundig model op basis van historische brandweerddata. *Tijdschrift voor Veiligheid*, 20(1), 60-74, doi.org/10.5553/TvV/.000020.
- Tannous, W.K. & K. Agho (2017) Socio-demographic predictors of residential fire and unwillingness to call the fire service in New South Wales. *Preventive Medicine Reports*, 7, 50-57, doi:10.1016/j.pmedr.2017.05.004.
- Turner, S.L., R.D. Johnson, A.L. Weightman, et al. (2017) Risk factors associated with unintentional house fire incidents, injuries and deaths in high-income countries: a systematic review. *Injury Prevention*, 23, 131-37.
- Verbond van Verzekeraars (2021) *Risicomonitor woningbranden*. www.verzekeraars.nl/publicaties/actueel/stijging-in-woningbrandclaims-2020, geraadpleegd 24 januari 2022.

Bijlage I Per buurtcluster de samenhangende attributen met coëfficiënt (volledige lijst van attributen in model 2)

attribuut	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5
(Intercept)	10,16	6,85	13,97	7,87	3,88
g_woz					0,11
p_lgezw				-1,03	
p_koopw		-0,21			
p_ov_hw	5,01				
p_e_o_w		0,36	0,11		0,08
p_bjj2k					0,3
p_hh_li					0,24
p_hh_lkk					0,21
p_hh_osm			2,93		0,35
g_pau_hh			-0,06		
g_pau_km	0,66			-0,02	
g_vernoo			-0,11		
p_45_64		-0,76			
p_gehuwd				-0,98	-0,27
p_gesch					0,33
p_w_all					0,53
p_marok					0,07
p_antaru	4,9		3,81		
p_suri				0,06	
p_tur	3,2		0,9		
p_ov_nw	0,81				-0,22
p_lp_hh			2,2		0,2
p_hh_z_k	-0,05	-0,63		-0,02	-0,53
p_soz_ww		-0,16			
p_bed_gi					0,02
p_bed_mn	-0,69				
OvermatigeDrinker_6					0,37
Overgewicht_8					-0,53
Roker_10			1,59		
ErnstigBeperktVanwegeGezondheid_13		0,28			
Gehoorbeperving_15			0,34		
Gezichtsbeperking_16		0,24			
Mobiliteitsbeperking_17		0,24			

(Vervolg)

attribuut	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5
HoogRisicoOpAngstOfDepressie_20				0,17	
Eenzaam_21				0,08	
Mantelzorger_29				-0,02	

Bijlage II Attributen die in de studie betrokken zijn

In de onderstaande tabel zijn alle attributen opgenomen die benut zijn voor de clusteranalyse van de buurten en tevens ook voor de Lasso-regressie. Dit zijn de overgebleven attributen na het opschonen en corrigeren van missende waardes. De uitleg over afkortingen en definities zijn te vinden in de referenties onder de tabel.

p_geb	p_ste	g_hhgro	g_woz	p_l_gezw
p_mgezw	p_bewndw	p_leegsw	p_koopw	p_huurw
p_wcorpw	p_ov_hw	p_e_o_w	p_bjj2k	p_bjo2k
g_ele	g_ele_ko	g_gas	g_gas_ko	g_ink_po
g_ink_pi	p_ink_li	p_ink_hi	p_n_act	p_hh_li
p_hh_hi	p_hh_lkk	p_hh_osm	g_pau_hh	g_pau_km
g_wodief	g_vernoo	g_gewsek	p_man	p_00_14
p_15_24	p_25_44	p_45_64	p_65_oo	p_ongeh
p_gehuwd	p_gesch	p_verwed	p_w_all	p_marok
p_antaru	p_suri	p_tur	p_ov_nw	p_lp_hh
p_hh_z_k	p_hh_m_k	p_soz_wb	p_soz_ao	p_soz_ww
p_soz_ow	p_bed_a	p_bed_bf	p_bed_gi	p_bed_hj
p_bed_kl	p_bed_mn	p_bed_ru	Drinker_4	ZwareDrinker_5
OvermatigeDrinker_6	VoldoetAanAlcoholrichtlijn_7	Overgewicht_8	Obesitas_9	Roker_10
LangdurigeZiekteOfAandoening_11	BeperktVanwegeGezondheid_12	ErnstigBeperktVanwegeGezondheid_13	LangdurigeZiekteEnBeperkt_14	Gehoorbepering_15
Gezichtsbeperking_16	Mobiliteitsbeperking_17	MinstensEenBeperking_18	MatigHoogRisicoOpAngstOfDepressie_19	HoogRisicoOpAngstOfDepressie_20
Eenzaam_21	ErnstigZeerErnstigEenzaam_22	GoedZeerGoedErvarenGezondheid_23	Sporter_24	VoldoetAanBeveegrichtlijn_25
MatigVeelRegieOverEigenLeven_26	MoeiteMetRondkomen_27	Vrijwilligerswerk_28	Mantelzorger_29	

Bron: www.cbs.nl/nl-nl/maatwerk/2016/30/kerncijfers-wijken-en-buurten-2016 en <https://data.overheid.nl/dataset/5209-gezondheid-per-wijk-en-buurt-2016#description>.