

**Te refereren<sup>1</sup> als:**

Snaphaan, T., Rummens, A., Van de Weghe, N., Van den Poel, D., Pauwels, L., & Hardyns, W. (2023). *Over hoe mobiele-telefoondata bijdragen aan de verbetering van criminaliteitsratio's en de voorspelling van criminaliteit in tijd en ruimte*. Vakgroep Criminologie, Strafrecht en Sociaal Recht, Universiteit Gent.

# Over hoe mobiele-telefoondata bijdragen aan de verbetering van criminaliteitsratio's en de voorspelling van criminaliteit in tijd en ruimte

Thom Snaphaan<sup>1,2\*</sup>, Anneleen Rummens<sup>1</sup>, Nico Van de Weghe<sup>3</sup>,  
Dirk Van den Poel<sup>4</sup>, Lieven Pauwels<sup>1</sup> & Wim Hardyns<sup>1,5</sup>

<sup>1</sup> Vakgroep Criminologie, Strafrecht en Sociaal Recht, Universiteit Gent, België.

<sup>2</sup> Centre of Expertise Veiligheid & Veerkracht, Avans Hogeschool, Nederland.

<sup>3</sup> Vakgroep Geografie, Universiteit Gent, België.

<sup>4</sup> Vakgroep Marketing, Innovatie en Organisatie, Universiteit Gent, België.

<sup>5</sup> Faculteit Sociale Wetenschappen, Universiteit Antwerpen, België.

\* Contact: [Thom.Snaphaan@UGent.be](mailto:Thom.Snaphaan@UGent.be)

## Samenvatting

In dit artikel wordt nagegaan of de *ambient population*, oftewel het aantal mensen dat op een bepaald tijdstip in een bepaald gebied aanwezig is, een geschiktere maatstaf biedt om het criminaliteitsrisico in tijd en ruimte in te schatten voor delicten met mobiele doelwitten dan de residentiële populatie. Meer specifiek wordt de potentie van het gebruik van deze *ambient population* als noemer in de berekening van criminaliteitsratio's en voorspeller van criminaliteitsrisico's voor predictieve modellen geëvalueerd. Hierbij wordt gebruikgemaakt van mobiele-telefoondata voor een periode van drie maanden in Gent ( $n = 595.858.852$  ruwe datapunten). Uit de resultaten blijkt dat de *ambient population* sterker samenhangt met geregistreeerde criminaliteit dan de residentiële populatie. Bovendien wijzen criminaliteitsratio's op basis van de *ambient population* ook andere probleemgebieden aan dan die op basis van de residentiële populatie. De voorspellingskracht van predictieve modellen voor tijdruimtelijke spreiding van

---

<sup>1</sup> Deze studie werd in het Engels, open access gepubliceerd in het peer-reviewed tijdschrift ISPRS International Journal of Geo-Information, zie: Rummens, A., Snaphaan, T., Van de Weghe, N., Van den Poel, D., Pauwels, L., & Hardyns, W. (2021). Do mobile phone data provide a better denominator in crime rates and improve spatiotemporal predictions of crime? *ISPRS International Journal of Geo-Information*, 10(6), Artikel 369. <https://doi.org/10.3390/ijgi10060369>

criminaliteit kan worden verbeterd door gebruik te maken van de *ambient population*. Deze resultaten bevestigen dat de *ambient population* een geschiktere maatstaf biedt voor delicten met een mobiel doelwit, omdat deze de onderliggende tijdruimtelijke dynamiek beter reflecteert.

### **Using mobile phone data to improve spatiotemporal crime rates and crime predictions**

This article assesses whether the ambient population is a more suitable population-at-risk measure for crime types with mobile targets than residential population. Specifically, the potential use of ambient population as a crime rate denominator and predictor for predictive policing models is evaluated, using mobile phone data ( $n = 595,858,852$  raw data points) as a proxy. The results show that ambient population correlates more strongly with crime than residential population. Crime rates based on ambient population designate different problem areas than crime rates based on residential population. The prediction performance of predictive policing models can be improved by using ambient population instead of residential population. These findings support that ambient population is a more suitable population-at-risk measure, as it better reflects the underlying dynamics in spatiotemporal crime trends.

## **1 Inleiding**

Onderzoek binnen de omgevingscriminologie leert dat criminaliteit significant varieert in tijd en ruimte, waarbij criminaliteitsconcentraties op specifieke tijdstippen (ook wel *burning times*) en specifieke plaatsen (ook wel *hotspots*) worden vastgesteld. (Tompson & Coupe, 2018; Weisburd, Bruinsma & Bernasco, 2009; Weisburd, Groff & Yang, 2012) De bestudeerde gebieden en tijdstippen variëren in verschillende opzichten, zoals qua omvang, bevolkingskenmerken en aantal bezoekers (bijv. werkgerelateerd of toeristen). Om met die verschillen rekening te houden, wordt binnen criminologisch onderzoek veelvuldig gebruikgemaakt van criminaliteitsratio's of -indices. Een criminaliteitsratio, ook wel *crime rate*, is een "statistic often used to represent the risk of criminal events [and that] help[s] to reveal clusters of crime in space and/or time based on an underlying population at risk" (Malleon & Andresen, 2015a, p. 112). Deze ratio's laten toe om een meer valide vergelijking te maken van verschillende ruimtelijke eenheden (bijv. een kleine stad versus een grootstad) en te controleren voor specifieke kenmerken van deze eenheden. Daarnaast reflecteren zij de risicopopulatie beter, waardoor betekenisvolle conclusies getrokken kunnen worden ten aanzien tijdruimtelijke patronen en de voorspellende of verklarende factoren daarvan. Een criminaliteitsratio wordt gecreëerd met behulp van een teller (aantal criminele incidenten) en een noemer (relevante risicopopulatie). Een veelgebruikte noemer in de berekening van criminaliteitsratio's is de residentiële populatie, oftewel de bevolking die op een bepaalde plaats woont. Data over deze residentiële populatie zijn relatief gemakkelijk te verkrijgen (veelal zijn deze openbaar beschikbaar) en er is aangetoond dat deze een sterke samenhang vertoont met criminaliteit in het algemeen. Om dezelfde reden wordt de residentiële populatie ook vaak gebruik als (controle)variabele in statistische modellen die gebruikt worden om tijdruimtelijke patronen van criminaliteit te voorspellen of verklaren.

Het gebruik van de residentiële populatie in criminaliteitsanalyses kent echter één belangrijk probleem: omdat het een statische maat is, wordt geen rekening gehouden met de tijdruimtelijke mobiliteit van daders, slachtoffers en *guardians* (Boggs, 1965; Pauwels, 2002). Dit komt bijvoorbeeld tot uiting in de dag- en nachtcycli, weekenden, vakanties en woon-werkverkeer (Stults & Hasbrouck, 2015, p. 346). Er is als het ware een eb en vloed van bezoekers (om diverse redenen) en bewoners die niet gereflecteerd wordt in een statische maat als de residentiële populatie. Zo kunnen ook onbewoonde gebieden met veel komen en gaan van mensen (bijv. parken en bedrijventerreinen) criminaliteit genereren of aantrekken (Brantingham & Brantingham, 1995). Bijgevolg biedt de residentiële populatie niet altijd, met name voor soorten criminaliteit met mobiele doelwitten en/of daders, een valide representatie van de daadwerkelijke (risico)populatie of potentiële doelwitten voor een bepaalde plaats en een bepaald tijdstip.

Een mogelijk alternatief voor deze residentiële populatie die deze tijdruimtelijke mobiliteit en dus de daadwerkelijke risicopopulatie beter kan representeren is de zogeheten *ambient population*. De *ambient population* betreft het aantal mensen dat op een bepaald tijdstip in een bepaald gebied aanwezig is (Andresen, 2006). De eerste initiatieven waarbij men gebruikmaakte van de *ambient population* dateren van midden jaren 2000, maar voornamelijk door de komst en alomtegenwoordigheid van smartphones en sociale media is het gebruik hiervan sinds circa 2014 aanzienlijk toegenomen. Dynamieken van individuen en mensenmassa's zijn in verband gebracht met criminaliteit en onderzoek wijst uit dat deze een significant effect hebben op de criminaliteitsratio's vanuit het idee dat "daily nonresidential activities distribute crime unevenly over space, beyond residential effects" (Felson & Boivin, 2014, p. 1). Het gebruik van de *ambient population* in plaats van de residentiële populatie kan daarom resulteren in een meer valide meting van de risicopopulatie en bijgevolg in betere toepassingen die zich baseren op deze maat, zoals criminaliteitsratio's en statistische modellen toegepast op tijdruimtelijke criminaliteitspatronen, zoals die gebruikt worden voor toepassingen als *predictive policing*.

In deze bijdrage presenteren we de resultaten van een studie waarbij we mobiele-telefoondata gebruiken als een proxymaat voor de *ambient population* in twee toepassingen: de analyse van criminaliteitsratio's en een predictieve analyse ten aanzien van het tijdruimtelijk criminaliteitsrisico. In beide gevallen zetten we de resultaten af tegen respectievelijk ratio's en predicties op basis van de residentiële populatie. Onze onderzoeksvraag is daarom: **In hoeverre is er een sterker verband tussen criminaliteit en de bevolkingsmaat wanneer de *ambient population* wordt gebruikt in vergelijking met de residentiële populatie?** Deze onderzoeksvraag valt uiteen in twee deelvragen:

1. In hoeverre verschillen de criminaliteitsratio's wanneer zij worden berekend op basis van de *ambient population* in vergelijking met de residentiële populatie?
2. Welke van de twee maten voor de risicopopulatie (*ambient population* en residentiële populatie) is een betere voorspeller in de predictieve analyse van tijdruimtelijke criminaliteitsrisico's?

We veronderstellen dat, zeker in geval van criminaliteitstypes met een mobiel doelwit, de *ambient population* een accuratere noemer biedt voor de berekening van criminaliteitsratio's en dat deze de predicties van tijdruimtelijke criminaliteitsrisico's verbetert.

## 2 Achtergrond

In deze sectie zullen we relevant voorgaand onderzoek uiteenzetten. Allereerst zullen we beknopt ingaan op twee voortdurende problemen in relatie tot het gebruik van de residentiële en *ambient population* in criminaliteitsanalyses. Ten tweede zullen we ingaan op de ontwikkelingen over de tijd heen, waarbij we verschillende generaties van databronnen onderscheiden die gebruikt worden als proxy van de *ambient population*. Ten derde zullen we voorgaand onderzoek samenvatten waarin gebruik werd gemaakt van mobiele-telefoondata om de *ambient population* te meten in relatie tot criminaliteit.

### 2.1 *Residentiële populatie versus ambient population: Gerelateerde uitdagingen*

Zoals eerder aangehaald, is het belangrijk om tijdruimtelijke variabiliteit van gerelateerde concepten in ogenschouw te nemen bij de bestudering van criminaliteit. Dit brengt twee onderling samenhangende uitdagingen met zich mee: enerzijds de keuze van (of de beschikbaarheid van data over) de meest geschikte (geldige) maat voor de risicopopulatie en anderzijds de keuze van (of beschikbaarheid van data over) de meest geschikte tijdruimtelijke eenheden van analyse.

#### 2.1.1 Bepaling van de meest geschikte maat voor de risicopopulatie

De risicopopulatie is een belangrijke maatstaf bij de berekening van criminaliteitsratio's en als (controle)variabele in tijdruimtelijke criminaliteitsmodellen. Boggs stelde in 1965 al dat het belangrijk is dat de noemers van criminaliteitsratio's gerelateerd moeten zijn aan de potentiële doelwitten of zoals Skogan (1976 in Solymosi & Bowers, 2018) het stelt aan "potential opportunities for victimization" (p. 216). In termen van de routineactiviteitentheorie (Cohen & Felson, 1979) moet de maat voor de risicopopulatie worden gerelateerd aan doelwitten van het specifieke type criminaliteit. Bijvoorbeeld in geval van criminaliteitsratio's voor autodiefstal bestaat de teller uit het aantal gestolen auto's in een bepaald gebied binnen een bepaalde periode en de noemer idealiter uit de som van alle auto's in datzelfde gebied in diezelfde periode, zodat de ratio het werkelijke risico reflecteert. Gegevens over de relevante risicopopulaties zijn echter niet altijd beschikbaar en daarom wordt vaak een proxy gebruikt om zo dicht mogelijk bij de werkelijke risicopopulatie in de buurt te komen (bijvoorbeeld het aantal parkeerplaatsen om het aantal auto's in een bepaald gebied te meten).

Het gebruik van de residentiële populatie gaat uit van de impliciete aanname dat dit een geschikte representatie is van de risicopopulatie, of in ieder geval een proxy van de werkelijke risicopopulatie. Uit voorgaand onderzoek blijkt echter dat het van belang is om te controleren voor doelwitten (zoals slachtoffers of objecten) die risico lopen en verschillen in gelegenheidsstructuren (Harries, 1991). Daarnaast is het belangrijk gebleken om de risicopopulatie te bepalen voor specifieke vormen van

criminaliteit (Boggs, 1965), omdat een geaggregeerde maat die meerdere vormen van criminaliteit omvat significante variatie kan verbergen of specifieke criminaliteitspatronen kan maskeren.

Het gebruik van de residentiële populatie als maat voor de risicopopulatie kan daarom problematisch zijn, aangezien voor bepaalde vormen van criminaliteit (1) personen niet het hoogste risico lopen om slachtoffer te worden van criminaliteit wanneer men thuis is, maar bijvoorbeeld wanneer men onderweg is (e.g., Lemieux & Felson, 2012). (2) Hoewel het een algemene tendens is dat daders criminaliteit plegen binnen hun *awareness spaces*, wordt niet alle criminaliteit gepleegd dichtbij hun woning, maar ook verder verwijderd van hun (huidige) woning, bijvoorbeeld in geval van overvallen (e.g., Bernasco, 2010; Groff & McEwen, 2007). (3) Mobiele doelwitten (zoals auto's of personen) bewegen in tijd en ruimte, en bijgevolg varieert ook het risico op slachtofferschap over tijd en ruimte. Het onderscheid tussen mobiele en immobiele doelen (Wikström, 1991) is met andere woorden essentieel in termen van het risico op slachtofferschap.

### 2.1.2 Bepaling van de meest geschikte analyse-eenheid

In geografisch onderzoek (niet uitsluitend gerelateerd aan de criminologie) wordt veel aandacht besteed aan het zogeheten *Modifiable Areal Unit Problem* (kortweg MAUP; Dark & Bram, 2007; Openshaw, 1984; Openshaw & Taylor, 1981; Parker, 1985), oftewel het probleem van de wijzigbare geografische eenheid. MAUP betreft het probleem van (keuzes in) de geografische aggregatie van gegevens en het daaruit voortvloeiende gevolg van welke gebeurtenissen tot dat gebied worden gerekend en welke niet. Dit probleem heeft twee dimensies: enerzijds de vraag welk zoneringsysteem (vorm of polygoon) van geografische eenheden het meest geschikt is in een specifieke studie (ook bekend als het *zonation effect* of zoneringseffect) en anderzijds de vraag wat het gevolg is van de gekozen mate van ruimtelijke (des)aggregatie of verandering van de geografische resolutie van de gegevens (ook bekend als het *scale effect* of schaaffect; Openshaw & Taylor, 1981). MAUP heeft veel aandacht gekregen in geografisch onderzoek binnen de criminologie (zie bijv. Gerell, 2017; Weisburd et al., 2009).

Studies binnen de omgevingscriminologie laten zien dat criminaliteit geconcentreerd is op microplaatsen (e.g., Groff et al., 2010; Weisburd et al., 2012), zelfs in die mate dat Weisburd een wet van criminaliteitsconcentratie op plaatsen postuleert die stelt dat “voor een gedefinieerde maat van criminaliteit op een specifieke microgeografische eenheid, zal de concentratie van criminaliteit vallen binnen een smalle bandbreedte van percentages voor een gedefinieerd cumulatief aandeel van criminaliteit” (Weisburd, 2015, p. 138). Hoewel er nog geen uniforme normen zijn vastgesteld voor het rapporteren en samenvatten van criminaliteitsconcentraties (Bernasco & Steenbeek, 2017) en deze ‘wet’ slechts een beschrijvend-empirische observatie betreft die meer vragen oproept dan antwoorden geeft (Hardyns et al., 2019), stellen Weisburd en collega's dat 50% van de criminaliteit geconcentreerd is op ongeveer 4% van de microplaatsen en 25% van de criminaliteit op minder dan 1,5% van de microplaatsen (Weisburd et al., 2016). Daarom moet bijzondere aandacht worden besteed aan het

onderzoek van criminaliteitstrends op kleine schaal, zoals adressen, fijnmazige rastercellen, straatsegmenten of clusters daarvan (Kounadi et al., 2018; Weisburd et al., 2009). Deze zogeheten *Criminology of Place* onderscheidt zich van eerdere denkrichtingen met belangstelling voor geografische aspecten van criminaliteit(preventie) door zich te richten op analyse-eenheden die kleiner zijn dan de census tracts of census block groups die frequent worden gebruikt om buurten te definiëren (Weisburd et al., 2012).

In tegenstelling tot dit geografisch probleem, is de temporele tegenhanger onderbelicht. Het *Modifiable Temporal Unit Problem* (kortweg MTUP) of het wijzigbare temporele eenheidsprobleem betreft het probleem van (keuzes in) gegevensaggregatie, maar dan op temporeel niveau (Cheng & Adepeju, 2014; Cöltekin et al., 2011). Dit bepaalt ook welke gebeurtenissen in een bepaald tijdsbestek worden opgenomen en welke niet. MTUP heeft drie dimensies: de eerste verwijst naar de temporele aggregatie of de eenheid van tijdwaarneming (*scale effect* of schaaffect, bijv. minuut, uur of dag); de tweede verwijst naar de manier waarop de temporele eenheden worden verdeeld (*segmentation effect* of segmentatie-effect, bijv. een week beginnen op zondag of maandag), en de derde verwijst naar de aanpassingen aan de temporele omvang van een tijdreeks (*boundary effect* of grenseffect, d.w.z. de arbitraire begin- en eindpunten van een tijdreeks; Cheng & Adepeju, 2014). Hoewel deze onwetendheid theoretisch moeilijk te rechtvaardigen is (van Sleeuwen et al., 2021), is zij verklaarbaar gezien de toenmalige onbeschikbaarheid van temporele fijnmazige gegevens. Momenteel zijn deze gegevensbeperkingen - theoretisch - niet langer van toepassing, aangezien nieuwe en opkomende databronnen (of big data) de belofte inhouden van “a data deluge – of rich, detailed, interrelated, timely and low-cost data – that can provide much more sophisticated, wider scale, finer grained understandings of societies and the world we live in” (Kitchin, 2013, p. 263). Deze nieuwe en opkomende databronnen, zoals GPS-gegevens en gegevens van mobiele telefoons, bieden grote mogelijkheden voor onderzoek in de omgevingscriminologie (Snaphaan & Hardyns, 2021a; Solymosi & Bowers, 2018).

Aangezien zowel de ruimtelijke als de temporele dimensie van belang zijn om de setting waarin regelovertredend gedrag plaatsvindt vast te leggen, is het aangewezen om de ‘eenheidsproblemen’ van (de convergentie van) beide dimensies te beoordelen. Dit resulteert in het *Modifiable Spatio-Temporal Unit Problem* (MSTUP) oftewel het “modificeerbare spatio-temporele eenheidsprobleem” (Martin et al., 2015). Beide dimensies variëren met de schaal, afhankelijk van de mate van zowel ruimtelijke als temporele heterogeniteit. Meentemeyer (1989) stelt dat “[i]n essence the scales need to match the heterogeneity; i.e., the phenomenon dictates the scale” (p. 171).

Met name de MSTUP vormt een hardnekkige uitdaging bij tijdruimtelijke analyses van criminaliteit. Eerder onderzoek laat zien dat niet alleen criminaliteit geconcentreerd is op microplaatsen en kleine tijdseenheden (e.g., Andresen & Malleson, 2015; Valente, 2019), maar ook dat een variërende tijdruimtelijke resolutie implicaties heeft voor de voorspelling van criminaliteit (Rummens & Hardyns,

2021). Het te klein maken van de schaal brengt specifieke problemen met zich mee, met als belangrijkste de beschikbaarheid en schaarsheid van de gegevens (d.w.z. een onevenredig groot aantal nullen). Dit wordt doorgaans veroorzaakt door de lage frequentie op een fijnmazige tijdruimtelijke schaal en vereist in het beste geval correctiemechanismen of in het slechtste geval een hoger aggregatieniveau om zinvolle conclusies te kunnen trekken (Curiel & Bishop, 2016; Mohler et al., 2019; Rummens & Hardyns, 2021).

## 2.2 *Ontwikkelingen in het meten van de ambient population*

In voorgaand onderzoek zijn verschillende methoden gebruikt om de *ambient population* te meten. We kunnen op basis hiervan verschillende ‘generaties’ van databronnen onderscheiden die worden gebruikt voor het meten (of beter: schatten) van de *ambient population*.

Ten eerste is het mogelijk om een maat te creëren voor de *ambient population* door de maat voor de residentiële populatie te corrigeren op basis van observaties of surveys, waarbij rekening wordt gehouden met het aantal inkomende en uitgaande pendelaars. Oberwittler (2004) gebruikte bijvoorbeeld het aantal reizigers in het openbaar vervoer als maatstaf voor de niet-residentiële populatie bij het schatten van de risicopopulatie. Later hebben Stults en Hasbrouck (2015) de variatie in criminaliteitsratio’s tussen steden bestudeerd, waarbij de instroom van forensen in de noemer van de criminaliteitsratio werd opgenomen. Felson en Boivin (2015) bestudeerden de variatie in criminaliteitsratio’s binnen een stad, waarbij zij gebruikmaakten van op enquêtes gebaseerde gegevens over de instroom van bezoekers om verschuivingen in de bevolking te meten en zo de omringende bevolking te schatten. Deze op enquêtes gebaseerde metingen leveren echter informatie op van een momentopname in de tijd.

Ten tweede zijn (aanvullende) administratieve data en *remote sensing* data (bijv. LandScan Global Population Database) gebruikt om de *ambient population* te schatten als alternatief voor de residentiële populatie (e.g., Andresen, 2006, 2007). Het LandScan-algoritme gebruikt ruimtelijke gegevens en beeldanalyse om bevolkingstellingsgegevens binnen een administratieve grens op te splitsen (Oak Ridge National Laboratory, z.d.). De LandScan-database bevat een gemiddelde bevolkingstelling voor gridcellen van ongeveer 1 km x 1 km en betreft een gemiddelde over 24 uur. Bij gebrek aan alternatieven heeft deze databron nuttige inzichten opgeleverd. Deze ruimtelijke en vooral temporele schalen zijn echter nog te grof om zinvol te zijn, aangezien ruimtelijke en temporele fluctuaties niet goed in beeld worden gebracht.

Ten derde is door gebruikers gegenereerde inhoud, zoals gegevens van sociale netwerksites, gebruikt als proxy-indicator voor de *ambient population*. Malleson en Andresen (2015a, 2015b, 2016) en Hipp et al. (2019) gebruikten Twitter-gegevens als een schatting van de *ambient population*. Zij gebruikten alle tweets met een geografische locatie in een bepaald gebied als maatstaf voor de *ambient population*. Kounadi et al. (2018) gebruikten een Twitter-dataset als aanvullende dataset om controlepunten te

extraheren om een bevolkingsdataset met grove resoluties te herverdelen over microplaatsen. Een andere databron afkomstig van socialenetwerksites die empirisch wordt gebruikt om op innovatieve wijze criminaliteitsratio's te berekenen is een FourSquare dataset. Kadar et al. (2017) gebruikten zowel plaatsen als check-ins van FourSquare bij het berekenen van maten voor de *ambient population*. Ondanks de mogelijkheden van deze zogeheten *user-generated* databronnen, zoals lage kosten en real-time beschikbaarheid, hebben deze databronnen ook tekortkomingen. In de eerste plaats hebben deze door gebruikers gegenereerde gegevens een grote dekkingsfout (Hsieh & Murphy, 2017; Snaphaan & Hardyns, 2021b), omdat de bevolking niet gelijk vertegenwoordigd is. Deze ongelijkheid is het gevolg van enerzijds de “digitale kloof” (Yu, 2006; d.w.z. de ongelijkheid in toegang tot technologie, die in dit geval uitgesproken wordt door het lagere gebruik van mobiele telefoons door specifieke groepen, zoals ouderen) en anderzijds het gebruik van geëtagde data die nodig zijn voor deze toepassing. Geëtagde data zijn een andere deelsteekproef binnen de deelsteekproef van gebruikers van sociale media (Malleon & Andresen, 2015b). Recentelijk hebben onderzoekers ook *user-generated* locatiegegevens over mobiele zendmasten gebruikt om een proxymaat van de *ambient population* te construeren (Johnson et al., 2020).

Ten vierde zijn mobiele-telefoondata gebruikt om de *ambient population* te schatten. Bogomolov en collega's (2014) kunnen worden gezien als pioniers op dit gebied binnen de criminologie. Zij vonden dat deze databron de nauwkeurigheid van voorspellingen van criminaliteit aanzienlijk verbetert, vergeleken met “traditional, rich – yet expensive to collect – statistical data about a borough's population” (Bogomolov et al., 2014, p. 433). We zullen in de volgende paragraaf verder ingaan op deze databron.

*2.3 Eerder onderzoek naar het gebruik van mobiele-telefoondata in de bestudering van criminaliteit*

Gegevens van mobiele telefoons bieden ongekende inzichten in menselijk gedrag en de dynamiek daarvan. In dit data-intensieve tijdperk creëren het wereldwijd gebruik van mobiele informatie- en communicatieapparaten en de menselijke interactie via deze mobiele apparaten digitale voetafdrukken van mensen, zowel in de ruimte als in de tijd (Järv et al., 2017; Kitchin, 2014). In 2018 gebruikte 96% van de Belgische inwoners tussen 16 en 74 jaar een mobiele telefoon, hetzij een smartphone (78%) en/of een gsm (20%) (Statbel, 2018). Deze alomtegenwoordige mobiele telefoons worden beschouwd als een proxymaat voor de *ambient population* (Ahas et al., 2010). Het is echter opmerkelijk dat dit aandeel niet gelijk verdeeld is over de algemene bevolking. Bijvoorbeeld, voor mensen in de jongste leeftijdscategorie (16-24 jaar) bedroeg het aandeel van het gebruik van een mobiele telefoon 99%, tegenover een aandeel van 80% voor mensen in de oudste leeftijdscategorie (65-74 jaar; merk op dat mensen onder de 16 jaar en boven de 74 jaar niet werden opgenomen in de studie van Statbel (2018)). Voor het gebruik van een smartphone was deze discrepantie nog groter: respectievelijk 95% versus 37% (Statbel, 2018).



In eerder empirisch onderzoek binnen de omgevingscriminologie is mobiele-telefoondata gebruikt als maat voor de *ambient population*. Deze eerdere studies (zie Tabel 1) hebben gemeen dat zij mobiele-telefoondata gebruikten om de dynamiek van personen in ruimte en tijd te meten, maar de methodologische aanpak verschilde tussen de studies. Alle studies vertrokken expliciet of impliciet van (aspecten van) gelegenheidstheorieën voor criminaliteit, meestal de routine-activiteitentheorie (Cohen & Felson, 1979) en/of de criminaliteitspatroontheorie (Brantingham & Brantingham, 2008). De meeste studies waren gericht op het voorspellen van tijdruimtelijke criminaliteitspatronen enerzijds en/of het verklaren van die tijdruimtelijke criminaliteitspatronen anderzijds. De mobiele-telefoondata werden uiteraard gebruikt als een proxymeting van de *ambient population*, maar de indicatoren die in de analyses werden opgenomen, waren afhankelijk van de beschikbaarheid van relevante uitsplitsingen in de dataset, zoals informatie over herkomst (i.e., het geschatte aantal inwoners, werknemers en bezoekers), geslacht (i.e., het geschatte aantal mannen en vrouwen) en leeftijdscategorieën (zie bijv. Bogomolov et al., 2014), waardoor het relatieve belang van de aanwezigheid van subgroepen binnen de *ambient population* kon worden onderzocht. Song en collega's (2019) verschilden in hun aanpak, omdat zij mobiele-telefoondata gebruikten om de mobiliteit van mensen te meten en louter de aanwezigheid van mensen in tijd en ruimte vaststelden. In tegenstelling tot de andere studies probeerden zij de locatiekeuze van daders te verklaren.

Wat de gebruikte methoden betreft, waren er diverse verschillen tussen de studies. Om te beginnen verschilden de studies, zoals samengevat in Tabel 1, aanzienlijk in de gebruikte ruimtelijke en temporele schalen. Ten tweede liepen de analytische benaderingen uiteen. Zoals eerder vermeld, probeerden sommige studies slechts tijdruimtelijke criminaliteitspatronen te voorspellen, terwijl andere probeerden (meer theoretisch onderbouwd) tijdruimtelijke variatie in criminaliteit te verklaren. Ten derde was er een verschil in de soorten criminaliteit die men in ogenschouw nam. De meeste studies waren gericht op diefstal van personen (He et al., 2020; Malleson & Andresen, 2016; Song et al., 2018; Song et al., 2019) of specifieke subtypes daarvan, zoals straatroof (Hanaoka, 2018). In de studies van Bogomolov et al. (2014, 2015) werd een algemene maat voor criminaliteit samengesteld op basis van elf verschillende criminaliteitstypen. Traunmueller et al. (2014) maakten een onderscheid tussen straatcriminaliteit (bijv. antisociaal gedrag, drugs, beroving en geweldsdelicten) en criminaliteit in en rond gebouwen (bijv. inbraak, vernieling en brandstichting, overige diefstal en winkeldiefstal). Haleem et al. (2020) en Lee et al. (2020) gebruikten in hun respectievelijke studies data over geweldsdelicten. Tabel 1 bevat informatie over verschillende algemene beschrijvende kenmerken van eerder onderzoek naar tijdruimtelijke patronen van criminaliteit met mobiele-telefoondata.

Alle relevante eerdere studies die zich richten op het voorspellen van criminaliteit vonden dat het invoeren van de maatstaf voor de *ambient population* een hogere predictieve accuraatheid opleverde dan de maatstaf voor de residentiële populatie (Bogomolov et al., 2014, 2015). Bovendien leverde het opnemen van de *ambient population* in alle relevante eerdere studies die gericht waren op het verklaren

van tijdruimtelijke variatie in criminaliteit, significante resultaten op in correlatieanalyses (Haleem et al., 2020; Malleson & Andresen, 2016; Traunmueller et al., 2014), en de toevoeging van deze maat in regressieanalyses verbeterde de modellen aanzienlijk (Hanaoka, 2018; He et al., 2020; Song et al., 2018). Twee belangrijke opmerkingen moeten echter worden gemaakt. Ten eerste leek de prestatie van de maatstaf voor de *ambient population* gedurende de dag te variëren, wat betekent dat op specifieke momenten van de dag andere proxy-maatstaven voor de *ambient population* beter presteerden (bijv. taxiriten in Song et al., 2018) of de maatstaf voor de werkdagbevolking (gebaseerd op een op enquête gebaseerde correctie van de dataset met betrekking tot de residentiële populatie; Malleson & Andresen, 2016)). Ten tweede moet worden opgemerkt dat alle conclusies van eerdere studies moeten worden gezien in het licht van hun eigen beperkingen en in het licht van de bestudeerde soorten criminaliteit.

Het is vermeldingswaardig om te benoemen dat mobiele-telefoondata op verschillende manieren (kunnen) worden benaderd (zie kolom Databron in Tabel 1). Men kan het activiteitsniveau gebruiken, zodat elke actie (bijv. sms'en of bellen) als een individueel geval in de dataset aanwezig is. Men kan ook het aantal unieke apparaten gebruiken, zodat elk apparaat (na ten minste één actie) slechts één keer wordt geteld. Bovendien kan men kiezen tussen verschillende typen van data die een telefoon genereert; men kan bijvoorbeeld elk type actie, records van gespreksgegevens of internetactiviteit opnemen. Ten slotte is een belangrijk nadeel van de meeste van de bovengenoemde studies dat de gebruikte databronnen in ruimte en/of tijd niet op elkaar aansluiten. Dit resulteert in een noodzakelijke aggregatie van andere gegevens omwille van de compatibiliteit en, erger nog, een gebrek aan fijnmazige tijdruimtelijke informatie. Zo zijn de meeste gebruikte criminaliteitsgegevens alleen op maandelijke basis beschikbaar, zonder variaties in de tijd (Bogomolov et al., 2014, 2015; He et al., 2020; Malleson & Andresen, 2016; Traunmueller et al., 2014), waardoor aggregatie van de fijnmazige (meestal per uur) mobiele-telefoondata noodzakelijk is (zie Tabel 1). Een ander voorbeeld, vanuit een ruimtelijk perspectief, is de studie van He et al. (2020), waarin ruimtelijk fijnmazige mobiele-telefoondata (gridcellen van 306 m x 306 meter) werden geaggregeerd tot aanzienlijk grotere analyse-eenheden (*Paichusuo*-gebieden; i.e., het werkgebied van een lokaal politiebureau; gemiddeld 54,34 km<sup>2</sup>).

**Tabel 1.** Algemene kenmerken van voorgaand onderzoek met betrekking tot tijdruimtelijke criminaliteitspatronen met behulp van mobiele-telefoondata.

Bron	Databron	Studiegebied	Ruimtelijke schaal	Temporele schaal
Bogomolov et al., 2014, 2015	Aantal unieke telefoongesprekken, geëxtrapoleerd naar de totale bevolking op basis van het marktaandeel van het netwerk in elke cel	Londen, VK	Onbekend; 124.119 cellen	Per uur, voor een periode van drie weken
Traunmueller et al., 2014	Personentelling (niet nader gespecificeerd)	Londen, VK	23.164 gridcellen van verschillende grootte (210 m x 210 m voor Inner London, 425 m x 425 m voor Outer London)	Per uur, voor een periode van drie weken
Malleson & Andresen, 2016	Mobiele-telefoonactiviteit	Londen, VK	4.835 <i>Lower Super Output Areas</i> (LSOA's) <sup>1</sup>	Per uur, voor een periode van een week
Hanaoka, 2018	Konzatsu Tokei ® data van mobiele telefoons met ingeschakelde auto-GPS-functie	Osaka City, Japan	Gridcellen van circa 250 m x 250 m	Per uur, voor een periode van twaalf maanden
Song et al., 2018	Cellulaire signaalgegevens: algemene 2G en 3G mobiele-telefoonactiviteit	“ZG City,” China (203 km <sup>2</sup> , >10.000.000 inwoners)	Gridcellen van 1 km x 1 km	Per uur, voor een periode van drie weken
Song et al., 2019	Cellulaire signaalgegevens: algemene 4G mobiele-telefoonactiviteit	“ZG City,” China (>3.000 km <sup>2</sup> , >5.000.000 inwoners)	1.616 census eenheden (gemiddeld oppervlak van 1,62 km <sup>2</sup> )	Per uur, voor een periode van één dag
He et al., 2020	Mobiele-telefoondata incl. informatie en activiteit van de gebruiker	Xi'an, China	Gridcellen van 306 m x 306 m	Per uur, voor een periode van vier maanden
Haleem et al., 2021; Lee et al., 2021	Mobiele telefoon herkomst-bestemming dataset	Manchester, VK	501 eenheden, verdeeld over 1.673 <i>Lower Super Output Areas</i> (LSOA's) <sup>1</sup>	17 tijdsvakken per uur en één tijdsvak tussen 23.00 en 05.59 uur, voor een periode van negentien dagen

<sup>1</sup> Lower Super Output Areas (LSOA's) zijn administratieve eenheden in het Verenigd Koninkrijk. Deze geografische eenheden hebben circa 1000 tot 3000 inwoners.

### 3 Data en methoden

In ons onderzoek omvat het studiegebied de stad en fusiegemeente Gent. Met 261.475 inwoners in 2018 is Gent na Antwerpen de grootste stad van België. In deze studie werden twee ruimtelijke analyse-eenheden gebruikt: het statistische-sectorniveau en het gridniveau. Statistische sectoren ( $n = 201$  voor Gent) zijn over het algemeen de kleinste betekenisvolle analyse-eenheden in België waarvoor systematisch demografische en sociaaleconomische gegevens worden verzameld en geanalyseerd. Ze zijn gebaseerd op sociaaleconomische, morfologische en ruimtelijke kenmerken (Statbel, z.d.). Ze

worden vaak gebruikt in sociaalecologisch onderzoek en beleids- en praktijktoepassingen (bijv. criminaliteitsstatistieken). Statistische sectoren staan ook wel bekend als buurten. Om de criminaliteitsratio's te vergelijken wanneer ze worden berekend met de *ambient population* versus de residentiële populatie (deelvraag 1), voerden wij onze analyse uit op het niveau van de statistische sector.

Bij de ruimtelijke modellering van criminaliteit worden in hedendaags onderzoek gewoonlijk gridcellen of straatsegmenten gebruikt, een nog kleiner niveau dan de statistische sector. Een belangrijke toepassing in de praktijk is de voorspelling van criminaliteitsincidenten en het gebruik ervan door politiediensten om patrouilles te optimaliseren (i.e., *predictive policing*). Deze toepassing vereist een micro-geografisch niveau om efficiënt te kunnen zijn. Daarom werd, om de *ambient population* en de residentiële populatie te vergelijken als voorspellers voor de voorspellende analyse van criminaliteit (deelvraag 2), een grid met een resolutie van 200 bij 200 meter gebruikt als ruimtelijk analyseniveau ( $n = 4206$  gridcellen voor Gent). Gridcellen met deze resolutie zijn ook eerder toegepast in empirisch criminologisch onderzoek (bijv. Hardyns et al., 2017; Hoeben et al., 2014; Rummens & Hardyns, 2021; Rummens et al., 2017).

### 3.1 Operationalisering en databronnen

De voor deze analyse gebruikte gegevens zijn afkomstig uit drie bronnen: criminaliteitsdata afkomstig van de Lokale Politie Gent, administratieve gegevens over de residentiële populatie afkomstig van de Stad Gent, en data over de *ambient population* afkomstig van de mobiele telefoonoperator Proximus.

De criminaliteitsdata die verkregen werden van de Lokale Politie Gent strekken van oktober tot en met december 2018. Data van drie afzonderlijke criminaliteitstypen werden gebruikt, namelijk: agressieve diefstal<sup>2</sup>, slagen en verwondingen<sup>3</sup>, en fietsdiefstal<sup>4</sup>. Daarbij moeten we de kanttekening maken dat de aangiftebereidheid (aan de kant van de burgers) en de registratiebereidheid (aan de kant van de politie) verschillen per criminaliteitstype. Zo is fietsdiefstal bijvoorbeeld een van de meest geregistreerde misdrijftypes in België en is het zelfs het meest geregistreerde type criminaliteit in Vlaanderen (Federale Politie, 2019a). Uit de resultaten van de meest recente Belgische Veiligheidsmonitor blijkt echter dat fietsdiefstal een van de criminaliteitstypen is met de laagste aangiftebereidheid van burgers: in de 12 maanden voorafgaand aan de dataverzameling gaf 10% van de Belgische huishoudens aan slachtoffer te zijn geweest van fietsdiefstal, maar slechts 48,1% meldde dit bij de politie (Federale Politie, 2019b).

We hebben ervoor gekozen om zowel gewelds- als vermogensdelicten mee te nemen in de analyse. Voor elk van de voornoemde criminaliteitstypen ontvingen we voor elke geregistreerde gebeurtenis in de onderzoeksperiode de locatie op adresniveau en het exacte tijdstip of de exacte tijdspanne waarin het

---

<sup>2</sup> Agressieve diefstal wordt gedefinieerd als handtasroof, diefstal met geweld en diefstal met vertoon of gebruik van een wapen, inclusief pogingen daartoe.

<sup>3</sup> Slagen en verwondingen wordt gedefinieerd als het opzettelijke gebruik van kracht of geweld dat resulteert in verwondingen, intrafamiliaal geweld is uitgesloten.

<sup>4</sup> Fietsdiefstal wordt gedefinieerd als eenvoudige diefstal van vergrendelde of ontgrendelde fietsen in de openbare ruimte, met inbegrip van pogingen daartoe.

delict vermoedelijk heeft plaatsgevonden (op basis van de informatie die het slachtoffer bij de registratie van het delict heeft verstrekt). Als alleen een tijdspanne beschikbaar was, werd ten behoeve van de analyse aangenomen dat het delict plaatsvond in het midden van die spanne. De criminaliteitsdata werden geocodeerd op basis van het Centraal referentieadressenbestand (CRAB) van Vlaanderen. Als er geen volledig adres (straat en huisnummer)<sup>5</sup> beschikbaar was, werd willekeurig een gridcel toegewezen uit de gridcellen die de straat overlaptten. De data waar geen straat werd geregistreerd<sup>6</sup>, werden uitgesloten van de analyse. Na opschoning van de data en geocodering bestaat de dataset uit in totaal 49 agressieve diefstaldelicten, 293 gevallen van slagen en verwondingen en 571 fietsdiefstallen over de volledige studieperiode.

De criminaliteitsdata werden vervolgens geaggregeerd tot het aantal delicten voor elk type criminaliteit en voor elke maand per statistische sector en per gridcel. Vanwege het grote aantal gridcellen met nul delicten (>95%) en het beperkte aantal cellen met meer dan één delict werd de variabele criminaliteit voor de analyse op gridniveau bijkomend gedichotomiseerd (i.e., 0 = geen geregistreerde delicten in de desbetreffende gridcel, in de desbetreffende maand en voor dat criminaliteitstype, 1 = één of meer geregistreerde delicten in de desbetreffende gridcel, in de desbetreffende maand en voor dat criminaliteitstype).

Data met betrekking tot de residentiële populatie werden verkregen via de Stad Gent. Zij verschaften het aantal inwoners per gridcel van 200 bij 200 meter van het door ons aangeleverde grid en hun respectievelijke statistische sectoren op basis van de op dat moment meest recente data uit het bevolkingsregister (i.e., 2018). Om privacyredenen maskeerde de Stad Gent gridcellen met vier of minder (maar niet gelijk aan nul) inwoners (i.e., 6,28% van de gridcellen). Die cellen werden getoond als cellen met vier inwoners. Een gridcel met vier inwoners betekent in onze dataset dus vier of minder inwoners. Gridcellen met nul inwoners (i.e., 48,19%) werden niet gemaskeerd.

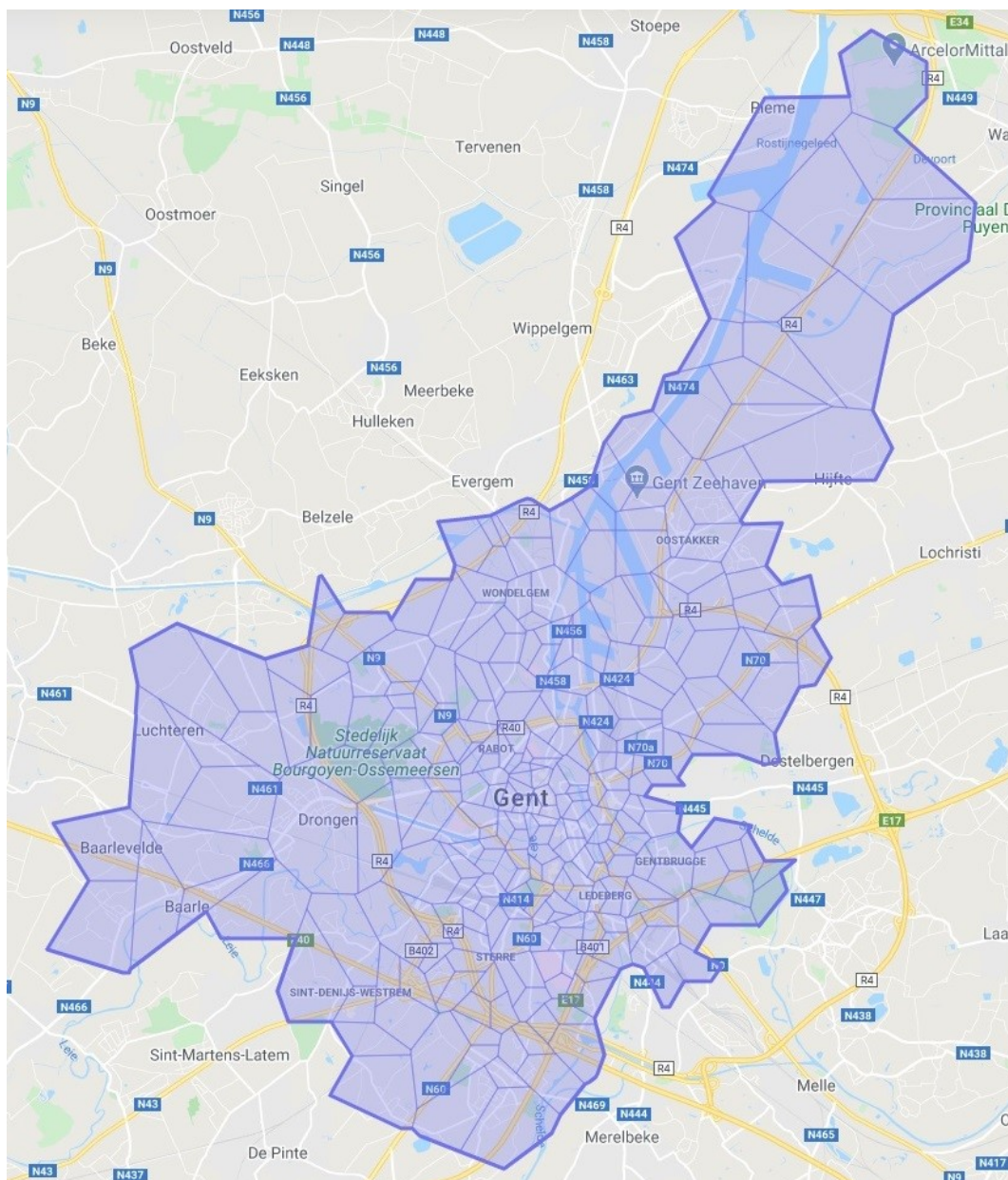
Ten slotte werd de *ambient population* geschat aan de hand van de mobiele-telefoondata van Proximus. Proximus is de grootste telecomoperator in België met een marktaandeel van 39,10% (Proximus, 2018). Meer bepaald bestaan de mobiele-telefoondata uit tellingen van individuele mobiele telefoons (unieke gebruikers) die aangesloten zijn op het Proximus-netwerk en aanwezig zijn in Gent. Om tellingen op een klein ruimtelijk niveau te produceren, hanteert Proximus een zeshoekig raster (Thiessen-polygonen) gecentreerd door hun mobiele zendmasten ( $n = 288$  cellen voor het gebied van Gent). De grootte van de individuele cellen hangt af van de bevolkingsdichtheid: hoe hoger de bevolkingsdichtheid, hoe kleiner de cellen, aangezien er in die gebieden meer zendmasten zijn (zie Figuur 1).

---

<sup>5</sup> Dit was voor de delicttypes agressieve diefstal, slagen en verwondingen, en fietsdiefstal in respectievelijk 67,80%, 35,06% en 34,15% van de delicten het geval.

<sup>6</sup> Dit was voor de delicttypes agressieve diefstal, slagen en verwondingen, en fietsdiefstal in respectievelijk 13,56%, 1,56% en 0,98% van de delicten het geval.

Het aantal aanwezige telefoons werd per uur geteld voor elke gridcel gedurende een periode van drie maanden. Dit resulteerde in een totaal van 9.397.473 datapunten in onze dataset, waarbij het gaat om 595.858.852 ruwe datapunten (i.e., som van het aantal unieke telefoons per gridcel per uur voor de periode oktober-december 2018). Om privacyredenen sloot Proximus cellen met 30 of minder aanwezige telefoons uit (1,06% van de datapunten in de ruwe gegevens; in de geaggregeerde datasets die voor onze analyse werden gebruikt, had geen enkele rij een *ambient population* van nul) en werden in totaal ‘slechts’ data van drie maanden verstrekt. De tellingen werden vervolgens geëxtrapoleerd naar de totale bevolking in verhouding tot het marktaandeel van Proximus in België. Ten slotte werden de tellingen in het zeshoekige raster (zie Figuur 1) omgezet naar de gridcellen van 200 bij 200 meter en toegewezen aan hun respectievelijke statistische sectoren.



**Figuur 1.** Het door Proximus gebruikte mobiele-telefoonnet in het studiegebied (Rummens et al., 2021).

### 3.2 Gebruikte methoden

Om het verband tussen criminaliteit en zowel de residentiële populatie als de *ambient population* in het algemeen te onderzoeken, werden correlatiecoëfficiënten berekend op zowel statistische-sectorniveau als op gridniveau. De correlatiecoëfficiënten werden berekend voor elk criminaliteitstype en voor elke maand afzonderlijk om te controleren op maandelijkse variaties. Op statistische-sectorniveau werd de Pearson-correlatiecoëfficiënt gebruikt en voor het gridniveau werd de punt-biseriële correlatiecoëfficiënt gebruikt vanwege de binaire aard van de criminaliteitsvariabele (zie hierboven). Tevens werd nagegaan of verschillen in correlaties tussen de residentiële populatie en *ambient population* met criminaliteit significant van elkaar verschilden met behulp van Zou's betrouwbaarheidsintervaltest (Zou, 2007; zie ook Diedenhofen & Musch, 2015).

Om na te gaan of het gebruik van de *ambient population* als noemer in de berekening van criminaliteitsratio's tot andere resultaten leidt dan het gebruik van de residentiële populatie als noemer (deelvraag 1), werden voor elke statistische sector criminaliteitsratio's berekend op basis van de residentiële populatie en de *ambient population*. Indien een statistische sector nul inwoners telde, werd deze sector uitgesloten, aangezien het in dat geval niet mogelijk zou zijn een criminaliteitsratio voor de residentiële populatie te berekenen. Slechts één sector werd om deze reden uitgesloten, en voor deze specifieke sector werden tijdens de studieperiode (oktober tot en met december 2018) geen criminaliteitsincidenten voor een van de drie criminaliteitstypen geregistreerd.

Om na te gaan of de *ambient population* een betere voorspeller voor criminaliteit biedt dan de residentiële populatie (deelvraag 2) gebruikten we predictieve analyses. Meer bepaald pasten we logistische-regressieanalyses toe om twee enkelvoudige regressiemodellen te bouwen: één met de residentiële populatie als voorspeller voor de kans dat een nieuw criminaliteitsincident zou plaatsvinden in elke gridcel en één met de *ambient population* als voorspeller voor die kans. De beschikbare data werden opgesplitst in een train- en een testset. De data van oktober en november werden gebruikt als de trainingsdataset om het voorspellingsmodel te trainen in het voorspellen van criminaliteitsincidenten. De data van december werden gebruikt als testset om de voorspellingsprestaties te evalueren. Met andere woorden, de locaties (gridcellen) waar criminaliteit plaatsvindt voor de maand december werden voorspeld. Om de modellen voor de woon- en *ambient population* te vergelijken, voorspelden beide modellen hetzelfde aantal criminaliteitsincidenten, afhankelijk van het gemiddelde aantal criminaliteitsincidenten per maand.

De predictieve prestaties van de modellen werden geëvalueerd aan de hand van de volgende maatstaven: trefkans, precisie, F1-score en AIC. Trefkans (EN: *recall*) is het aandeel correct voorspelde incidenten ten opzichte van het totale aantal incidenten. Precisie (EN: *precision*) is het percentage correct voorspelde gridcellen ten opzichte van het totale aantal gridcellen met een risico. Idealiter heeft een goed predictief model zowel een hoge *trefkans* als *hoge precisie*. Om dit weer te geven hebben we ook

de F1-score opgenomen. De F1-score is het harmonisch gemiddelde van *trekkan* en *precisie* en beschouwt beide dus in één maatstaf. Het *Akaike Informatie Criterium* (AIC) is een maatstaf waarmee verschillende modellen op een eenduidige wijze kunnen worden vergeleken, rekening houdend met zowel de *goodness-of-fit* als de eenvoud van het model. De AIC schat de relatieve hoeveelheid informatie die door het model verloren gaat: hoe minder informatieverlies, hoe beter het model. Het model met de laagste AIC-waarde is daarom doorgaans het betere model. De traditioneel veelgebruikte ROC-curve en de bijbehorende oppervlakte onder de curve (EN: AUC) maat werden in deze studie niet gebruikt, omdat zij misleidend kunnen zijn wanneer er sprake is van een matige tot ernstige onevenwichtigheid tussen de klassen (Davis & Goadrich, 2006; Saito & Rehmsmeier, 2015), zoals hier het geval was vanwege de relatief lage frequentie van criminaliteit.

## 4 Resultaten

Om meer inzicht te krijgen in het verband tussen enerzijds de woon- en *ambient population* en anderzijds criminaliteit werd in deze studie allereerst gebruikgemaakt van correlatieve analyses. Ten tweede werd in kaart gebracht in welke mate de criminaliteitsratio's verschilden naargelang de woon- of *ambient population* werd gebruikt als noemer in de berekening ervan (deelvraag 1). Ten derde werd middels predictieve analyses nagegaan of de woon- of *ambient population* een betere voorspeller was voor het tijdrumtelijk criminaliteitsrisico. De resultaten worden in de volgende paragrafen opeenvolgend beschreven.

### 4.1 Correlatieve analyse

Tabel 2 bevat de correlatiecoëfficiënten op het niveau van de statistische sector voor elk van de drie criminaliteitstypen voor de woon- versus *ambient population* voor elke maand die bestudeerd werd. In alle gevallen was het verband tussen criminaliteit en bevolking, zowel de woon- als de *ambient population*, positief. De correlatie tussen criminaliteit en bevolking was significant met  $p < 0.001$  voor alle criminaliteitstypen. Voor alle drie de criminaliteitstypen werd een sterkere correlatie vastgesteld met de *ambient population* dan met de residentiële populatie. Het sterkste verband was er tussen de *ambient population* en fietsdiefstal (met een correlatiecoëfficiënt van 0.60 in november). De correlatieverschillen waren ook significant, met uitzondering van agressieve diefstal in oktober en december, waar geen significant verschil werd vastgesteld. In het algemeen was het verband tussen criminaliteit en bevolking consistent tussen de maanden, met uitzondering van agressieve diefstal (met name residentiële populatie), wat verklaard zou kunnen worden door de relatief lagere frequentie van dat delicttype.



**Tabel 2.** Pearson correlatiecoëfficiënten van criminaliteit per delicttype en per maand, afgezet tegen de woon- en *ambient population* (statistische-sectorniveau,  $n=201$ ).

Delicttype	Maand	Residentiële populatie	<i>Ambient population</i>	Correlatieverschil
Agressieve diefstal	okt	0.24***	0.36***	0.12
	nov	0.15*	0.35***	0.20*
	dec	0.26***	0.23***	0.03
Slagen en verwondingen	okt	0.19**	0.46***	0.27*
	nov	0.18*	0.44***	0.26*
	dec	0.23**	0.41***	0.18*
Fietsdiefstal	okt	0.30***	0.56***	0.26*
	nov	0.31***	0.60***	0.29*
	dec	0.26***	0.55***	0.29*

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

Wanneer de correlaties op het gridniveau worden bekeken (zie Tabel 3), blijkt dat ook in dit geval de *ambient population* een sterker verband vertoont met criminaliteit dan de residentiële populatie. De verschillen zijn echter minder uitgesproken dan bij de analyse op het statistische-sectorniveau, wat waarschijnlijk te wijten is aan de veel kleinere ruimtelijke schaal en het daaruit voortvloeiende beperkt aantal observaties per eenheid. Alle correlatiecoëfficiënten zijn positief en significant. Het sterkste verband stellen we opnieuw vast tussen de *ambient population* en fietsdiefstal (met een correlatiecoëfficiënt van 0.47 in oktober). Alle correlatieverschillen zijn significant, wat bevestigt dat de *ambient population* een significant sterkere correlatie had met criminaliteit voor alle delicttypes dan de residentiële populatie.

**Tabel 3.** Punt-biseriële correlatiecoëfficiënten van criminaliteit per delicttype en per maand, afgezet tegen de woon- en *ambient population* (gridniveau,  $n=4206$ ).

Delicttype	Maand	Residentiële populatie	<i>Ambient population</i>	Correlatieverschil
Agressieve diefstal	okt	0.12***	0.16***	0.04*
	nov	0.08***	0.18***	0.10*
	dec	0.12***	0.13***	0.01*
Slagen en verwondingen	okt	0.23***	0.26***	0.03*
	nov	0.23***	0.25***	0.02*
	dec	0.21***	0.22***	0.01*
Fietsdiefstal	okt	0.34***	0.47***	0.13*
	nov	0.27***	0.41***	0.14*
	dec	0.23***	0.35***	0.12*

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

#### 4.2 Criminaliteitsratio's

Een belangrijke deelvraag in deze studie is of het gebruik van de *ambient population* in plaats van de residentiële populatie de berekening van criminaliteitsratio's ten goede komt. Om deze reden werd het aantal delicten voor de verschillende delicttypes (teller) afgezet tegen respectievelijk de woon- en *ambient population* (noemer). Door de teller tegen de noemer af te zetten, wordt een (criminaliteits)ratio verkregen. Voor de meeste statistische sectoren kwamen de uitkomsten gebaseerd op berekeningen met de *ambient population* en de residentiële populatie met elkaar overeen. Er waren echter verschillende

sectoren waar het gebruik van de *ambient population* of de residentiële populatie als noemer voor de criminaliteitsratio's een significant verschil maakte. In Tabel 4 worden deze sectoren, waarbij een relatief groot verschil bestond tussen de criminaliteitsratio's gebaseerd op de woon- en *ambient population* (i.e., met een verschil van >2 tussen de gestandaardiseerde scores), inzichtelijk gemaakt voor elk delicttype. Hierbij zijn ook belangrijke kenmerken en nabijgelegen oriëntatiepunten opgenomen.

**Table 4.** Statistische sectoren met verschillende criminaliteitsratio's op basis van de woon- en *ambient population*, en diens belangrijke kenmerken.

<b>Sector ID</b>	<b>Kenmerken en oriëntatiepunten in de buurt</b>	<b>Criminaliteitsratio op basis van residentiële populatie (gestandaardiseerd)</b>	<b>Criminaliteitsratio op basis van <i>ambient population</i> (gestandaardiseerd)</b>
<b>Agressieve diefstal</b>			
C72 Muidebrug	Verhoogd armoederisico, gebied met veel verkeer	0,228	3,617
<b>Slagen en verwondingen</b>			
A00 Kuip	Stadscentrum, studentenbuurt, hoge concentratie van bars, restaurants en winkels	5,613	2,979
A321 Sint-Pieters	Uitgaansgebied, studentenbuurt	6,704	4,622
A46 Blaarmeersen	Natuur-, sport- en recreatiedomein	7,541	0,149
A542 Groendreef	Park, politiebureau	0,358	3,578
B452 Sint-Alois	Concentratie van scholen	6,705	4,428
B472 Groothandelsmarkt	Voetbalstadium (Ghelamco Arena), dichtbij ziekenhuis	8,536	0,079
C72 Muidebrug	Verhoogd armoederisico, gebied met veel verkeer	0,228	3,617
C772 Vormingsstation-Oost	Treindepot, dichtbij groot treinstation	5,232	0,072
J172 Bugten	Evenementenhal (Flanders Expo)	3,815	0,512
J197 Maria Middelaers	Ziekenhuis, dichtbij evenementenhal (Flanders Expo)	3,453	0,334
K622 Heilig Huizeken	Dichtbij natuurgebied (Hoge Lake)	0,145	2,156
<b>Fietsdiefstal</b>			
A00 Kuip	Stadscentrum, studentenbuurt, hoge concentratie van bars, restaurants en winkels	11,796	6,514
A35 Station	Groot treinstation	4,574	8,074
A45 Groene vallei	Park, dichtbij gevangenis en politiebureau	0,909	8,922
A46 Blaarmeersen	Natuur-, sport- en recreatiedomein	4,936	-0,029
A50 Drongensesteenweg	Knooppunt van meerdere hoofdwegen	0,410	2,480
A542 Groendreef	Park	0,147	2,256
E32 Dampoort	Groot treinstation	1,793	6,500
K613 Oude Wee	Sporthal, voetbalveld en golfclub	0,392	3,306
K022 Oude Abdij	Klein treinstation	2,032	8,294

Voor deze sectoren gaven de criminaliteitsratio's op basis van de *ambient population* doorgaans een ander beeld dan de criminaliteitsratio's op basis van de residentiële populatie. In sommige gevallen was de criminaliteitsratio op basis van de *ambient population* hoger dan die op basis van de residentiële populatie (bijv. sector C72), terwijl ze in andere gevallen lager was (bijv. sector A46). In het laatste geval waren de verschillen grotendeels te wijten aan de verschillen in omvang tussen de *ambient population* en de residentiële populatie. De criminaliteitsratio's op basis van de *ambient population* waren meestal lager in sectoren met faciliteiten die veel mensen aantrekken (bijv. treinstation, sportstadion, recreatiegebied) in relatief dunbevolkte gebieden. Dit wijst erop dat, hoewel de

criminaliteit relatief hoog lijkt wanneer wordt gekeken naar de residentiële populatie, deze niet zo hoog is wanneer rekening wordt gehouden met het werkelijke aantal mensen dat in dat gebied komt en gaat. Deze gebieden kunnen worden beschouwd als *crime generators*, oftewel knooppunten waar dagelijkse activiteiten plaatsvinden en die een groot aantal personen aantrekken, waaronder potentiële doelwitten en daders (Brantingham & Brantingham, 1995). In het geval van fietsdiefstal is de situatie in sector A46 opmerkelijk, aangezien de criminaliteitsratio op basis van de *ambient population* lager dan gemiddeld was voor deze sector, terwijl de criminaliteitsratio op basis van de residentiële populatie aanzienlijk hoger dan gemiddeld was.

Interessanter zijn de sectoren waar de situatie omgekeerd was, i.e. waar de criminaliteitsratio's op basis van de *ambient population* gemiddeld hoger was. In veel van deze sectoren gebeurde dit ondanks de eerder genoemde discrepantie tussen een hoge *ambient population* en een lage residentiële populatie (zie bijv. de sectoren A35 en E32 die gelegen zijn in de nabijheid van grote treinstations). Deze gebieden kunnen beschouwd worden als *crime attractors*, oftewel plaatsen of voorzieningen die de reputatie hebben dat er gelegenheden voor criminaliteit aanwezig zijn en waar daders naartoe trekken met de bedoeling om criminaliteit te plegen (Brantingham & Brantingham, 1995). Dit betekent dat zelfs wanneer rekening wordt gehouden met de feitelijk aanwezige bevolking de concentratie van criminaliteit in die sectoren hoog is. In andere sectoren (bijv. de sectoren C72 en A542) is het niet onmiddellijk duidelijk wat de verklaring is van de relatief hogere criminaliteit. Het gaat in ieder geval om buurten met een hoge criminaliteitsconcentratie die onopgemerkt waren gebleven indien criminaliteitsratio's berekend werden op basis van de residentiële populatie.

#### 4.3 *Predictieve analyse*

De tweede deelvraag van deze studie heeft betrekking op in welke mate de prestaties van tijdruimtelijke voorspellingen van criminaliteit verbeteren naargelang de woon- of *ambient population* wordt gebruikt als voorspeller. Gebaseerd op data van de eerste twee maanden (oktober-november 2018) maakten we voorspellingen van criminaliteitsincidenten voor agressieve diefstal, slagen en verwondingen, en fietsdiefstal in december 2018. Zoals vermeld in de paragraaf 'Data en methoden', betreffen dit enkelvoudige regressiemodellen om de predictieve prestaties van de woon- en *ambient population* in isolatie na te gaan. Tabel 5 bevat de resulterende predictieve prestaties van beide modellen voor ieder delicttype afzonderlijk, wat een vergelijking toelaat.

**Tabel 5.** Prestaties van de predictieve modellen.

	Treffkans	Precisie	F1-score	AIC
<b>Agressieve diefstal</b> ( <i>n</i> delicten = 20, <i>n</i> predicties = 20)				
Model met residentiële populatie	5,00%	4,00%	0,044	281
Model met <i>ambient population</i>	25,00%	8,00%	0,121	246
<b>Slagen en verwondingen</b> ( <i>n</i> delicten = 97, <i>n</i> predicties = 100)				
Model met residentiële populatie	10,31%	8,00%	0,090	1206
Model met <i>ambient population</i>	40,21%	10,00%	0,160	1198
<b>Fietsdiefstal</b> ( <i>n</i> delicten = 100, <i>n</i> predicties = 150)				
Model met residentiële populatie	20,00%	10,67%	0,139	2002
Model met <i>ambient population</i>	61,00%	22,67%	0,386	1718

Voor agressieve diefstal werden ‘slechts’ 20 misdrijven geregistreerd – vanuit maatschappelijk oogpunt is lager natuurlijk beter, maar vanuit methodologisch oogpunt levert een laag absoluut aantal in veel gevallen problemen op. Het model met de residentiële populatie kon slechts één incident voorspellen. Het model met de *ambient population* daarentegen presteerde beter door 25,00% (vijf criminaliteitsincidenten) te voorspellen met een precisie van 8,00%, wat betekent dat van de 20 voorspelde gridcellen met een hoog risico, in twee cellen daadwerkelijk een of meer delicten werd(en) gepleegd in december 2018. Kijkend naar de predictieve prestaties voor slagen en verwondingen, stellen we vast dat het model met de *ambient population* (met een treffkans van 40,21%) beter presteert dan het model met de residentiële populatie (treffkans van 10,31%), hoewel de precisie in beide gevallen relatief laag was (respectievelijk 8,00% en 10,00%). In het geval van fietsdiefstal behaalden zowel het model met de residentiële populatie als het model met de *ambient population* de hoogste prestatiescores van alle drie de delicttypes, maar het model met de *ambient population* presteerde opnieuw beter dan het model met de residentiële populatie. Het model met de *ambient population* was zelfs in staat een relatief groot percentage delicten (61,00%) te voorspellen met de hoogste precisiesscore (22,67%), ook al was het de enige variabele in het model.

Kortom, uit alle prestatiemetingen blijkt dat het model met de *ambient population* beter presteert dan het model met de residentiële populatie voor alle bestudeerde criminaliteitstypes. Er werd vooral een groot verschil geconstateerd op vlak van de treffkans. Dit lijkt erop te wijzen dat de modellen met de *ambient population* vooral goed zijn in het voorspellen van locaties met sterke concentraties van criminaliteit. In het geval van fietsdiefstal bijvoorbeeld waren de incidenten sterk geconcentreerd bij openbare fietsenstallingen (bijv. bij de treinstations).

## 5 Discussie en conclusie

Het gebruik van de *ambient population* in plaats van de residentiële populatie voor de analyse van criminaliteitsratio's biedt verschillende voordelen. Ten eerste is de meting van de *ambient population* dynamischer dan die van de residentiële populatie, waardoor een meer dynamische analyse van de criminaliteit mogelijk is, waarbij bijvoorbeeld rekening wordt gehouden met maandelijkse en seizoensgebonden variaties. Ten tweede geeft de *ambient population* ook beter de bevolking weer die

op een bepaald tijdstip en een bepaalde plaats werkelijk aanwezig is, wat vooral belangrijk is in gebieden met weinig inwoners maar veel bezoekers, zoals winkelgebieden. Ten derde voorkomt het gebruik van de *ambient population* mogelijk gegevensverlies in gebieden waar wel criminaliteit is maar geen residentiële populatie, wat vooral een probleem is bij het gebruik van kleine geografische eenheden. Ten slotte biedt de *ambient population* voor de bestudeerde soorten criminaliteit een betere afspiegeling van de risicobevolking. Mobiele-telefoondata vormen een nauwkeurige benadering van de *ambient population*, aangezien het gebruik van mobiele telefoons tegenwoordig wijdverbreid is en deze telefoons overal worden meegenomen. Bovendien kunnen met mobiele-telefoondata ook de *ambient population* in real-time worden gemeten, wat de mogelijkheid biedt om nog meer voordeel te halen uit het dynamische karakter ervan. In onze analyse vertoonde de *ambient population* een sterker verband met agressieve diefstal, slagen en verwondingen, en fietsdiefstal dan de residentiële populatie. Er moet echter worden opgemerkt dat de robuustheid van onze resultaten met de nodige voorzichtigheid moet worden geïnterpreteerd, aangezien we voor onze studie slechts data voor drie maanden konden verkrijgen, wat ook betekent dat we slechts een beperkt aantal delicten konden bestuderen en dat de conclusies met betrekking tot mogelijke seizoensgebonden effecten beperkt zijn. Niettemin ondersteunen onze bevindingen die van eerder onderzoek en kunnen zij aanknopingspunten bieden voor toekomstig onderzoek.

Deze studie wijst uit dat criminaliteitsratio's berekend op basis van de *ambient population* in sommige gevallen andere gebieden uitlichtte dan wanneer criminaliteitsratio's berekend worden op basis van de residentiële populatie. Dit heeft implicaties voor beleid dat gevormd wordt door of preventie-initiatieven die genomen worden ten aanzien van probleemgebieden. Bovendien leidde het gebruik van de *ambient population* als voorspeller in plaats van de residentiële populatie tot correctere voorspellingen van criminaliteit. Het opnemen van de *ambient population* in plaats van de residentiële populatie als variabele zou toepassingen van criminaliteitsmodellering, zoals *predictive policing* toepassingen, kunnen verbeteren, vooral omdat het model met de *ambient population* beter presteert op vlak van het voorspellen van locaties met hoge criminaliteitsconcentraties, wat in het licht van de *law of crime concentration at places* (zie Hardyns et al., 2017) een zeer wenselijke eigenschap is.

Ondanks de potentiële voordelen zijn er ook enkele uitdagingen die zich voordoen bij het gebruik van mobiele-telefoondata als proxy voor de *ambient population*. Hoewel een van de belangrijkste voordelen het dynamische karakter is, leidt dit ook tot een gecompliceerder dataverzamelingsproces met verschillende praktische uitdagingen. Hoewel smartphones alomtegenwoordig zijn, is het duidelijk dat het gebruik van mobiele telefoons niet gelijk verdeeld is over de populatie. Deze 'digitale kloof' (*digital divide*), in dit geval het lagere mobiele-telefoongebruik door specifieke groepen (zoals ouderen), beïnvloedt de representativiteit van mobiele-telefoondata als proxy voor de *ambient population*. Niettemin is deze kloof met betrekking tot de leeftijd van personen die de *ambient population* uitmaken minder problematisch dan op het eerste gezicht lijkt, gezien de consistente bevinding binnen de

ontwikkelingscriminologie met betrekking tot de sterke relatie tussen leeftijd en het plegen van criminaliteit (ook bekend als de *age-crime curve*; zie bijv. Farrington, 1989), de empirische bevinding dat ouderen ook minder slachtoffer zijn dan jongere leeftijdscategorieën (zie bijv. Morgan & Oudekerk, 2018), en het feit dat smartphones nog steeds alomtegenwoordig zijn in de oudere leeftijdsgroepen (zo gebruikt 80% van de Belgische inwoners tussen 65 en 74 jaar een mobiele telefoon). De belangrijkste beperking van mobiele-telefoondata heeft echter te maken met de onnauwkeurigheid van de locatieschattingen van individuele toestellen (Malleon & Andresen, 2016). Een meer praktisch probleem is de moeilijkheid om mobiele-telefoondata te verkrijgen voor wetenschappelijke onderzoeksdoeleinden omwille van verschillende factoren, zoals de bereidheid van telecomoperatoren om de data ter beschikking te stellen, beperkingen opgelegd door de Algemene Verordening Gegevensbescherming (bijv. de limiet van drie maanden in onze studie), en mogelijke privacyproblemen. Een andere factor waarmee rekening moet worden gehouden is het marktaandeel van de telecomoperator: hoe groter dit is, hoe representatiever de gegevens. Extrapolatie naar de totale bevolking is immers slechts een beperkte oplossing, aangezien het in wezen gaat om een schatting van het werkelijke aantal. Ten slotte, hoewel de *ambient population* geschikter is voor bepaalde soorten criminaliteit, is voor andere delicttypes, zoals woninginbraak, eerder de residentiële populatie de meest geschikte operationalisering van de risicobevolking.

Gezien de mogelijke verschillen tussen delicttypes wat betreft het gebruik van *ambient population* in plaats van residentiële populatie (als gevolg van het onderscheid tussen mobiele en immobiele doelwitten, en de specificiteit daarvan), moeten criminaliteitstypes steeds afzonderlijk worden bestudeerd, in plaats van te kijken naar ‘de criminaliteit’ als algemene maatstaf. Het voornoemde indachtig, dient toekomstig onderzoek de analyse uit deze studie uit te breiden naar andere criminaliteitstypen en andere studiegebieden, om na te gaan of deze observaties ook in andere contexten gelden. Daarnaast zou het interessant zijn de verklarende mechanismen tussen de *ambient population* en criminaliteit nader te onderzoeken (Wikström & Treiber, 2016). In deze studie ligt de nadruk op de geschiktheid van de met behulp van mobiele-telefoondata gemeten *ambient population* voor het schatten van criminaliteitsratio's en met het oog op het voorspellen van criminaliteit. Toekomstig onderzoek zou het effect van deze voorspellende variabele kunnen beoordelen in combinatie met andere relevante variabelen die mogelijk bijdragen aan criminaliteitsconcentraties op microplaatsen (bijv. kenmerken van landgebruik; e.g. Caplan & Kennedy, 2016; Wheeler & Steenbeek, 2020). Ten slotte bieden nieuwe en opkomende databronnen en innovatieve dataverwerkingsmethoden voortdurend nieuwe mogelijkheden voor wetenschappelijk onderzoek. Een integratie met andere proxies voor de *ambient population* (zoals Wi-Fi-gegevens, zie Crols & Malleon, 2019) of andere potentiële big-databronnen (bijv. data van commerciële bedrijven) zou ook nader kunnen worden onderzocht. Dankzij innovatieve dataverwerkingsmethoden (e.g. convolutionele neurale netwerken, zie bijv. Gebru et al., 2017) kunnen onderzoekers en praktijkmensen data verwerken die omvangrijker en gevarieerder zijn, en met een hoge

snelheid geproduceerd worden (Zhang et al., 2018). Deze methoden bieden ook mogelijkheden om de meest geschikte noemer voor de berekening van criminaliteitsratio's geldig te meten (bijv. het tellen van het aantal fietsen in geografische gebieden door middel van *computer vision* om te gebruiken als noemer in de berekening van de criminaliteitsratio voor fietsendiefstal).

Deze inspanningen met betrekking tot het gebruik van nieuwe en opkomende databronnen zijn essentieel voor de ontwikkeling van de meest geschikte en nauwkeurige noemer bij de berekening van criminaliteitsratio's en andere risicometingen. De resultaten hebben implicaties voor zowel het criminologisch onderzoek als voor beleid en praktijk. Met deze metingen kunnen onderzoekers de mechanismen ontleden die verband houden met bijvoorbeeld de tijdruimtelijke convergentie van personen met een criminele geneigdheid en criminogene omgevingen (Wikström, 2014) of de tijdruimtelijke verspreiding van geschikte doelwitten. Een belangrijke piste voor toekomstig onderzoek, aangezien we weten dat mobiele-telefoondata als databron een geldig alternatief vormen voor bestaande metingen, is het optimaliseren van de databron, in overeenstemming met de nieuwste theoretische inzichten. We weten immers dat het belangrijk is om de samenstelling van de *ambient population* mee te nemen en in kaart te hebben welke activiteiten zij verrichten (Wikström et al., 2012). Voorgaand onderzoek toont aan dat het mogelijk is om onderscheid te maken naargelang bepaalde kenmerken van de *ambient population* in mobiele-telefoondata (Bogomolov et al., 2014; Malleson & Andresen, 2016; Traunmueller et al., 2014). Zo zou het bijvoorbeeld nauwkeuriger zijn om rekening te houden met de aan risico's 'blootgestelde populatie', in plaats van de *ambient population* (Haleem et al., 2020). De uitdaging ligt in het verkrijgen van gegevens die het mogelijk maken dit onderscheid te maken en dit onderscheid (methodologisch en theoretisch correct) te gebruiken om geïntegreerde criminologische theorieën te testen. Toekomstig onderzoek moet uitwijzen welke operationalisering van de risicopopulatie het meest geschikt is en onder welke omstandigheden (bijv. voor welk specifiek type criminaliteit, verschillen in tijdstip).

Tegelijkertijd heeft het probleem van het kiezen van een geschikte noemer voor criminaliteitsratio's ook belangrijke gevolgen voor het beleid en de praktijk. Beleidsmakers maken rijkelijk gebruik van criminaliteitsratio's om hun beleid vast te stellen, te verdedigen en te evalueren, maar het gebruik van de juiste noemer kan een verschil maken bij het vaststellen van probleemgebieden in termen van criminaliteitsratio's. Ook politie en andere wetshandhavers die gebruikmaken van analyse van criminaliteitsdata en voorspellende toepassingen om proactief patrouilles uit te voeren, zouden er baat bij kunnen hebben om rekening te houden met een adequatere risicopopulatie. Gelet op de verbetering van de voorspellingsprestaties bij het gebruik van de *ambient population*, is het gebruik van deze variabele vooral van belang voor *predictive policing* modellen, aangezien zij tijdruimtelijke fluctuaties op microniveau kunnen weergeven en derhalve nauwkeuriger voorspellingen op deze kleine schaal mogelijk maken. Dit is volledig in lijn met één van de strategische doelstellingen in het Nationaal Veiligheidsplan 2022-2025, waarin wordt gesteld dat "de politie [inzet] op het ontwikkelen en inzetten

van tools en procedures die de gegevensstroom kunnen capteren en filteren, zodat een bruikbare informatiebron ontstaat die kan worden aangewend voor politieopdrachten en -onderzoeken” (Politie, z.d., p. 24). Deze studie laat zien hoe mobiele-telefoondata zinvol ingezet kunnen worden in de beeldvorming over criminaliteitsfenomenen. Toekomstig onderzoek en toepassingen van predicties van criminaliteitsrisico’s die volledig gebruikmaken van de dynamische aard van de *ambient population* dienen echter een daadwerkelijke voorspellende analytische strategie toe te passen, die de *ambient population* op een eerder tijdstip includeert als voorspeller van criminaliteit.

Concluderend stellen wij dat de *ambient population* een betere afspiegeling is van de risicopopulatie en de relevante mechanismen beter weergeeft (bijv. wat betreft de aard van het doelwit: mobiel of immobiel) en daarom veel potentieel heeft in criminologisch onderzoek, theorietoetsing, beleid en praktijk. Mobiele-telefoondata zijn in dit verband een veelbelovende proxy, aangezien zij een grote hoeveelheid data op een fijnmazige tijdruimtelijke schaal kunnen opleveren (voor onze studie in totaal 595.858.852 ruwe datapunten in een tijdspanne van drie maanden). In deze studie hebben wij deze potentie empirisch bevestigd en zijn wij een van de eersten die het gebruik ervan voorstellen en onderzoeken als alternatief voor het gebruik van de veelgebruikte residentiële populatie voor het berekenen van criminaliteitsratio’s en het voorspellen van criminaliteitsrisico’s.



## Referenties

- Ahas, R., Silm, S., Järvi, O., Saluveer, E., & Tiru, M. (2010). Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of Urban Technology*, 17(1), 3-27. <https://doi.org/10.1080/10630731003597306>
- Andresen, M. A. (2006). Crime measures and the spatial analysis of criminal activity. *British Journal of Criminology*, 46(2), 258-285. <https://doi.org/10.1093/bjc/azi054>
- Andresen, M. A. (2007). Location quotients, ambient populations, and the spatial analysis of crime in Vancouver, Canada. *Environment and Planning A: Economy and Space*, 39(10), 2423-2444. <https://doi.org/10.1068/a38187>
- Andresen, M. A., & Malleson, N. (2015). Intra-week spatial-temporal crime patterns. *Crime Science*, 4, Article 12. <https://doi.org/10.1186/s40163-015-0024-7>
- Bernasco, W. (2010). A sentimental journey to crime: Effects of residential history on crime location choice. *Criminology*, 48, 389-416.
- Bernasco, W., & Steenbeek, W. (2017). More places than crimes: Implications for evaluating the law of crime concentration at place. *Journal of Quantitative Criminology*, 33(3), 451-467. <https://doi.org/10.1007/s10940-016-9324-7>
- Boggs, S. L. (1965). Urban crime patterns. *American Sociological Review*, 30(6), 899-908.
- Bogomolov, A., Lepri, B., Staiano, J., Letouzé, E., Oliver, N., Pianesi, F., & Pentland, A. S. (2015). Moves on the street: Classifying crime hotspots using aggregated anonymized data on people dynamics. *Big Data*, 3(3), 148-158. <https://doi.org/10.1089/big.2014.0054>
- Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. S. (2014). *Once upon a crime: Towards crime prediction from demographics and mobile data*. Paper presented at the 16<sup>th</sup> International Conference on Multimodal Interaction, Istanbul, Turkey (pp. 427-434). <https://doi.org/10.1145/2663204.2663254>
- Brantingham, P. J., & Brantingham, P. L. (2008). Crime pattern theory. In R. Wortley & L. Mazarolle (Eds.), *Environmental criminology and crime analysis* (pp. 78-94). Willan Publishing.
- Brantingham, P. L., & Brantingham, P. J. (1995). Criminality of place. Crime generators and crime attractors. *European Journal on Criminal Policy and Research*, 3(3), 5-26.
- Caplan, J. M., & Kennedy, L. W. (2016). *Risk terrain modelling: Crime prediction and risk reduction*. University of California Press.
- Cheng, T., & Adepeju, M. (2014). Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PloS one*, 9(6), Article e100465. <https://doi.org/10.1371/journal.pone.0100465>
- Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44(4), 588-608. <https://doi.org/10.2307/2094589>
- Cöltekin, A., De Sabbata, S., Willi, C., Vontobel, I., Pfister, S., Kuhn, M., & Lacayo, M. (2011). *Modifiable temporal unit problem*. ISPRS/ICA workshop on Persistent Problems in Geographic

- Visualization (ICC2011). International Cartographic Association.  
<http://geoanalytics.net/ica/icc2011/coltekin.pdf>
- Crols, T., & Malleson, N. (2019). Quantifying the ambient population using hourly population footfall data and an agent-based model of daily mobility. *GeoInformatica*, 23, <https://doi.org/10.1007/s10707-019-00346-1>
- Curiel, R. P., & Bishop, S. (2016). A measure of the concentration of rare events. *Scientific Reports*, 6, Article 32369. <https://doi.org/10.1038/srep32369>
- Dark S.J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography: Earth and Environment*, 31(5), 471-479. <https://doi.org/10.1177/0309133307083294>
- Davis, J. & Goadrich, M. (2006). The relationship between precision-recall and the ROC curves. In W. Cohen & A. Moore (Eds.), *Proceedings of the 23<sup>rd</sup> International Conference on Machine Learning* (pp. 233–240). Association for Computing Machinery.
- Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE*, 10, Article e0121945. <https://doi.org/10.1371/journal.pone.0121945>
- Farrington, D. P. (1986). Age and crime. In M. Tonry & N. Morris (Eds.), *Crime and justice: An annual review of research* (Vol. 7, pp. 189–250). University of Chicago Press.
- Federale Politie. (2019a). *Criminaliteitsstatistieken*.  
<http://www.stat.policefederale.be/criminaliteitsstatistieken/interactief/>
- Federale Politie. (2019b). *Veiligheidsmonitor 2018. Grote tendensen*.  
[http://www.moniteurdeseurite.policefederale.be/assets/pdf/2018/reports/Grote\\_tendensen\\_Analyses\\_VMS2018.pdf](http://www.moniteurdeseurite.policefederale.be/assets/pdf/2018/reports/Grote_tendensen_Analyses_VMS2018.pdf)
- Felson, M., & Boivin, R. (2015). Daily crime flows within a city. *Crime Science*, 4, Article 31. <https://doi.org/10.1186/s40163-015-0039-0>
- Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Lieberman Aiden, E., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, 114(50), 13108-13113.
- Gerell, M. (2017). Smallest is better? The spatial distribution of arson and the Modifiable Areal Unit Problem. *Journal of Quantitative Criminology*, 33, 293-318. <https://doi.org/10.1007/s10940-016-9297-6>
- Groff, E. R., & McEwen, T. (2007). Integrating distance into mobility triangle typologies. *Social Science Computer Review*, 25, 210–238.
- Groff, E. R., Weisburd, D., & Yang, S.-M. (2010). Is it important to examine crime trends at a local “micro” level?: A longitudinal analysis of street to street variability in crime trajectories. *Journal of Quantitative Criminology*, 26, 7-32. <https://doi.org/10.1007/s10940-009-9081-y>

- Haleem, M. S., Lee, W. D., Ellison, M., & Bannister, J. (2021). The ‘exposed’ population, violent crime in public space and the night-time economy in Manchester, United Kingdom. *European Journal on Criminal Policy and Research*, 27, 335–352. <https://doi.org/10.1007/s10610-020-09452-5>
- Hanaoka, K. (2018). New insights on relationships between street crimes and ambient population: Use of hourly population data estimated from mobile phone users’ locations. *Environment and Planning B: Urban Analytics and City Science*, 45(2), 295-311.
- Hardyns, W., Snapphaan, T., & Pauwels, L. (2017). Criminaliteitsconcentraties en microplaatsen. Een toets van de ‘law of crime concentration at places’. *Tijdschrift voor Criminologie*, 59(3), 227–244. <https://doi.org/10.5553/tvc/0165182x2017059003001>
- Hardyns, W., Snapphaan, T., & Pauwels, L.J.R. (2019). Crime concentrations and micro places: An empirical test of the “law of crime concentration at places” in Belgium. *Australian & New Zealand Journal of Criminology*, 52(3), 390-410. <https://doi.org/10.1177/0004865818807243>
- Harries, K. D. (1991). Alternative denominators in conventional crime rates. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental Criminology* (pp. 147–165). Waveland Press.
- He, L., Páez, A., Jiao, J., An, P., Lu, C., Mao, W., & Long, D. (2020). Ambient population and larceny-theft, A spatial analysis using mobile phone data. *ISPRS International Journal of Geo-Information*, 9(6), Article 342. <https://doi.org/10.3390/ijgi9060342>
- Hipp, J. R., Bates, C., Lichman, M., & Smyth, P. (2019). Using social media to measure temporal ambient population: Does it help explain local crime rates? *Justice Quarterly*, 36(4), 718-748. <https://doi.org/10.1080/07418825.2018.1445276>
- Hoeben, E.M., Bernasco, W., Weerman, F.M., Pauwels, L., & van Halem, S. (2014). The space-time budget method in criminological research. *Crime Science*, 3(1), Article 12. <https://doi.org/10.1186/s40163-014-0012-3>
- Hsieh, Y. P., & Murphy, J. (2017). Total Twitter Error: Decomposing public opinion measurement on Twitter from a Total Survey Error perspective. In P.P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, & B. T. West (Eds.), *Total Survey Error in Practice* (pp. 23-46). John Wiley & Sons.
- Järv, O., Tenkanen, H., & Toivonen, T. (2017). Enhancing spatial accuracy of mobile phone data using multi-temporal dasymetric interpolation. *International Journal of Geographical Information Science*, 31(8), 1630-1651.
- Johnson, P., Andresen, M. A., Malleson, N. (2020). Cell towers and the ambient population: A spatial analysis of disaggregated property crime. *European Journal on Criminal Policy and Research*. <https://doi.org/10.1007/s10610-020-09446-3>
- Kadar, C., Brüngger, R. R., & Pletikosa, I. (2017). Measuring ambient population from location-based social networks to describe urban crime. In G. L. Ciampaglia, A. Mashhadi & T. Yasseri (Eds.), *Social Informatics* (pp. 521-535). [https://doi.org/10.1007/978-3-319-67217-5\\_31](https://doi.org/10.1007/978-3-319-67217-5_31)

- Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), 262-267.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage Publications.
- Kounadi, O., Ristea, A., Leitner, M., & Langford, C. (2018). Population at risk: Using areal interpolation and Twitter messages to create population models for burglaries and robberies. *Cartography and Geographic Information Science*, 45(3), 205-220.
- Lee, W., Haleem, M. S., Ellison, M., & Bannister, J. (2021). The influence of intra-daily activities and settings upon weekday violent crime in public spaces in Manchester, UK. *European Journal on Criminal Policy and Research*, 27, 375 – 395. <https://doi.org/10.1007/s10610-020-09456-1>
- Lemieux, A.M., & Felson, M. (2012). Risk of violent crime victimization during major daily activities. *Violence and Victims*, 27(5), 635-655.
- Malleson, N., & Andresen, M. A. (2015a). The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science*, 42(2), 112-121. <https://doi.org/10.1080/15230406.2014.905756>
- Malleson, N., & Andresen, M. A. (2015b). Spatiotemporal crime hotspots and the ambient population. *Crime Science*, 4, Article 10. <http://dx.doi.org/10.1186/s40163-015-0023-8>
- Malleson, N., & Andresen, M. A. (2016). Exploring the impact of ambient population measures on London crime hotspots. *Journal of Criminal Justice*, 46, 42-63. doi: 10.1016/j.jcrimjus.2016.03.002
- Martin, D., Cockings, S., & Leung, S. (2015). Developing a flexible framework for spatiotemporal population modeling. *Annals of the Association of American Geographers*, 105(4), 754-772.
- Meentemeyer, V. (1989). Geographical perspectives of space, time, and scale. *Landscape Ecology*, 3, 163-173. <https://doi.org/10.1007/BF00131535>
- Mohler, G., Brantingham, P. J., Carter, J., & Short, M. B. (2019). Reducing bias in estimates for the law of crime concentration. *Journal of Quantitative Criminology*, 35, 747-765. <https://doi.org/10.1007/s10940-019-09404-1>
- Morgan, R. E., & Oudekerk, B. A. (2019, September). *Criminal victimizations, 2018*. U.S. Department of Justice. <https://www.bjs.gov/content/pub/pdf/cv18.pdf>
- Oak Ridge National Laboratory. (n.d.). *Documentation*. Retrieved from <https://landscan.ornl.gov/documentation/#inputData>
- Oberwittler, D. (2004, September). Re-balancing routine activity and social disorganization theories in the explanation of urban violence. A new approach to the analysis of spatial crime patterns based on population at risk. *Social problems and juvenile delinquency in ecological perspective* (Working paper, no. 10). [https://pure.mpg.de/rest/items/item\\_2501318/component/file\\_3021556/content](https://pure.mpg.de/rest/items/item_2501318/component/file_3021556/content)

- Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and Planning A: Economy and Space*, 16(1), 17–31.
- Openshaw, S., & Taylor, P.J. (1981). The modifiable areal unit problem. In N. Wrigley & R. J. Bennett (Eds.), *Quantitative Geography: A British View* (pp. 60-70). Routledge and Kegan Paul.
- Parker, R. (1985). Aggregation, ratio variables, and measurement problems in criminological research, *Journal of Quantitative Criminology*, 1, 269-280. <https://doi.org/10.1007/BF01064636>
- Pauwels, L. (2002). *De ene buurt is de andere niet: Exploratie van mogelijkheden tot contextualisering van geregistreerde criminaliteit op buurtniveau*. VUB Press.
- Politie. (z.d.). *Nationaal Veiligheidsplan 2022-2025. Een veerkrachtige Geïntegreerde Politie, ten dienste van de samenleving*. <https://www.politie.be/5998/sites/5998/files/files/2022-04/NVP2022-2025.pdf>
- Proximus. (2018). *Financiële resultaten van de Proximus Groep – Eerste kwartaal 2018*. <https://www.proximus.com/nl/news/financial-results-q1-2018>
- Ridgeway, G. (2018). Policing in the era of big data. *Annual Review of Criminology*, 1 401-419. <https://doi.org/10.1146/annurev-criminol-062217-114209>
- Rummens, A., & Hardyns, W. (2021). The effect of spatiotemporal resolution on predictive policing model performance. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2020.03.006>
- Rummens, A., Hardyns, W. & Pauwels, L. (2017). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. *Applied Geography* 86, 255-261. <https://doi.org/10.1016/j.apgeog.2017.06.011>
- Rummens, A., Snaphaan, T., Van de Weghe, N., Van den Poel, D., Pauwels, L., & Hardyns, W. (2021). Do mobile phone data provide a better denominator in crime rates and improve spatiotemporal predictions of crime? *ISPRS International Journal of Geo-Information*, 10(6), Artikel 369. <https://doi.org/10.3390/ijgi10060369>
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, 10(3), Article e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Snaphaan, T., & Hardyns, W. (2021a). Environmental criminology in the big data era. *European Journal of Criminology*, 18(5), 713–734. <https://doi.org/10.1177/1477370819877753>
- Snaphaan, T., & Hardyns, W. (2021b). Handvatten voor een kwaliteitsbeoordeling van big data : de introductie van het Total Error raamwerk. *Tijdschrift voor Veiligheid*, 20(4), 63–88. <https://doi.org/10.5553/TvV/.000033>
- Solymosi, R., & Bowers, K. (2018). The role of innovative data collection methods in advancing criminological understanding. In G.J.N. Bruinsma & S. D. Johnson (Eds.), *The Oxford Handbook of Environmental Criminology* (pp. 210-237). Oxford University Press.

- Song, G., Bernasco, W., Liu, L., Xiao, L., Zhou, S., & Liao, W. (2019). Crime feeds on legal activities: Daily mobility flows help to explain thieves' target location choices. *Journal of Quantitative Criminology*, 35, 831-854. <http://doi.org/10.1007/s10940-019-09406-z>
- Song, G., Liu, L., Bernasco, W., Xiao, L., Zhou, S., & Liao, W. (2018). Testing indicators of risk populations for theft from the person across space and time: The significance of mobility and outdoor activity. *Annals of the American Association of Geographers*, 108(5), 1370-1388. <https://doi.org/10.1080/24694452.2017.1414580>
- Statbel. (2018). *ICT-gebruik in huishoudens*. [https://statbel.fgov.be/sites/default/files/files/documents/Huishoudens/10.5%20ICT-gebruik%20in%20huishoudens/TabIn2018\\_NI\\_2019-03-29.xlsx](https://statbel.fgov.be/sites/default/files/files/documents/Huishoudens/10.5%20ICT-gebruik%20in%20huishoudens/TabIn2018_NI_2019-03-29.xlsx)
- Statbel. (z.d.). *Statistische sectoren*. <https://statbel.fgov.be/nl/overstatbel/methodologie/classificaties/statistische-sectoren>
- Stults, B. J., & Hasbrouck, M. (2015). The effect of commuting on city-level crime rates. *Journal of Quantitative Criminology*, 31, 331–350. <https://doi.org/10.1007/s10940-015-9251-z>
- Taylor, B., Kowalyk, A., Boba, R. (2007). The integration of crime analysis into law enforcement agencies: An exploratory study into the perceptions of crime analysts. *Police Quarterly*, 10(2), 154-169. <https://doi.org/10.1177/1098611107299393>
- Tompson, L., & Coupe T. (2018). Time and opportunity. In G.J.N. Bruinsma & S.D. Johnson (Eds.), *The Oxford handbook of environmental criminology* (pp. 695–719). Oxford University Press.
- Traunmueller, M., Quattrone, G., & Capra, L. (2014). Mining mobile phone data to investigate urban crime theories at scale. In L. M. Aiello & D. McFarland (Eds.), *Social Informatics* (pp. 396-411). Springer.
- Valente, R. (2019). Spatial and temporal patterns of violent crime in a Brazilian state capital: A quantitative analysis focusing on micro places and small units of time. *Applied Geography*, 103, 90-97.
- van Sleeuwen, S. E. M., Ruiters, S., & Steenbeek, W. (2021). Right place, right time? Making crime pattern theory time-specific. *Crime Science*, 10, Article 2. <https://doi.org/10.1186/s40163-021-00139-8>
- Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology*, 53(2), 133-157. <https://doi.org/10.1111/1745-9125.12070>
- Weisburd, D., Bruinsma, G.J.N., & Bernasco, W. (2009). Units of analysis in geographic criminology: Historical development, critical issues, and open questions. in Weisburd, D. Bernasco, W. & Bruinsma, G.J.N. (Eds.) *Putting crime in its place* (pp. 3-31). Springer.
- Weisburd, D., Eck, J. E., Braga, A. A., Telep, C. W., Cave, B., Bowers, K., Bruinsma, G., Gill, C., Groff, E. R., Hibdon, J., Hinkle, J. C., Johnson, S. D., Lawton, B., Lum, C., Ratcliffe, J. H., Rengert, G., Taniguchi, T., & Yang, S.-M. (2016). *Place matters: Criminology for the twenty-first century*. Cambridge University Press.

- Weisburd, D., Groff, E.R. & Yang, S.-M. (2012). *The criminology of place: street segments and our understanding of the crime problem*. Oxford University Press.
- Wheeler, A. P., & Steenbeek, W. (2021). Mapping the risk terrain for crime using machine learning. *Journal of Quantitative Criminology*, 37, 445–480. <https://doi.org/10.1007/s10940-020-09457-7>
- Wikström P.-O. (2014). Situational Action Theory. In G. Bruinsma & D. Weisburd (Eds.) *Encyclopedia of Criminology and Criminal Justice*. Springer.
- Wikström, P.-O. (1991). *Urban crime, criminals and victims: The Swedish experience in an Anglo-American comparative perspective*. Springer-Verlag.
- Wikström, P.-O., & Treiber, K. (2016). Situational theory: The importance of interactions and action mechanisms in the explanation of crime. In A. Piquero (Ed.), *Handbook of Criminological Theory* (pp. 414-444). Wiley-Blackwell.
- Wikstrom, P.-O., Oberwittler, D., Treiber, K., & Hardie, B. (2012). *Breaking rules: The social and situational dynamics of young people's urban culture*. Oxford University Press.
- Yu, L. (2006). Understanding information inequality: Making sense of the literature of the information and digital divides. *Journal of Librarianship and Information Science*, 38(4), 229-252. <https://doi.org/10.1177/0961000606070600>
- Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, 42, 146-157. <https://doi.org/10.1016/j.inffus.2017.10.006>
- Zou G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, 12(4), 399–413. <https://doi.org/10.1037/1082-989X.12.4.399>