



**Center for
Big Data Statistics**

Working paper

Online vacatures als bron voor statistiek

Combineren van informatie uit online vacatures met
Vacature Enquête

Working paper no.: 04-20

Martine de Mooij
Dominik Blatt
Chantal Melser
Marko Roos

Maart 2020

Inhoudsopgave

1	Online vacatures als bron voor statistiek	3
2	Methodologie	3
2.1	Bron	3
2.2	Bewerking vacaturedata	5
2.3	Weging	10
3	Rol intermediaire vacatures in het weegmodel	13
3.1	Verschillen weegmethodiek UWV en CBS	14
3.2	Invloed aanpassing volgorde weegstappen	15
3.3	Conclusie methode CBS	18
4	Resultaten	20
4.1	Beroepsklassen	20
4.2	Vacatures Zorg & welzijn naar branche en naar regio	20

1 Online vacatures als bron voor statistiek

Het aantal nieuwe vacatures is een belangrijke indicator voor het verloop van de Nederlandse conjunctuur. Daarnaast is het van belang te weten in welke bedrijfstakken en functies de spanning op de arbeidsmarkt het grootst is, en of gevraagde vaardigheden aansluiten op de vaardigheden van mensen die nieuw op de arbeidsmarkt komen. Het CBS onderzoekt de mogelijkheid om met online vacatures een vacaturestatistiek te ontwikkelen, die achtergrondinformatie over vacatures weergeeft met publicatie met meer detail van bedrijfstakken. Deze nieuwe statistiek zou een aanvulling zijn op de bestaande [Vacature Enquête \(VE\)](#).

Teksten van online vacatures bevatten onder meer informatie over de functie, het gevraagde opleidingsniveau, de gewenste competenties, de werkgever en de werklocatie. Wanneer het CBS deze gegevens uit online vacatures kan gebruiken, hoeft deze informatie niet uitgevraagd te worden bij bedrijven.

Ieder jaar worden in Nederland miljoenen vacatureteksten geplaatst op websites van bedrijven en intermediairs. Deze vacatures kunnen verzameld worden via webscraping. Vervolgens zijn ze niet zonder meer bruikbaar als bron voor statistiek, omdat ze geen representatief beeld geven van alle nieuwe vacatures op de Nederlandse arbeidsmarkt. Een deel van de vacatures verschijnt niet op internet, terwijl een groot deel juist meerdere keren op internet verschijnt. Soms is dat zowel bij het bedrijf dat personeel vraagt (directe vacature) als bij een intermediair (indirecte vacature). Daarom is een methode nodig om de gewebscrapete nieuwe vacatures te wegen naar een representatief aantal nieuw ontstane vacatures op de Nederlandse arbeidsmarkt.

In dit rapport wordt een methode gepresenteerd om online vacatures te wegen. Het CBS heeft daarbij voortgeborduurd op een weegmodel dat ontwikkeld is door het Uitvoeringsinstituut Werknemersverzekeringen (UWV). Door het model toe te passen worden de Nederlandse vacatures afgeleid naar onder andere beroep. Het CBS stelt een aantal aanpassingen in de methode voor, met name in de manier waarop intermediaire vacatures worden opgenomen in het model.

Dit artikel bestaat uit drie onderdelen. In Hoofdstuk 2 wordt het weegmodel nader toegelicht. In Hoofdstuk 3 wordt inzichtelijk gemaakt welke invloed intermediaire vacatures hebben op het weegmodel. In Hoofdstuk 4 worden vacatures per beroepsklasse en hoofdbranche gepresenteerd volgens het nieuwe weegmodel.

2 Methodologie

2.1 Bron

Online vacatures via webscraping

Online vacatures worden verzameld via webscraping. Hierbij wordt eerst systematisch gezocht naar nieuwe advertenties (postings) op websites. Dit kunnen zowel websites zijn van bedrijven zelf, maar ook van jobboards (vacaturebanken). Vervolgens worden de teksten van de advertenties in het geheel gekopieerd. Dit proces is uitgevoerd door het bedrijf Textkernel, dat gespecialiseerd is in semantische zoek- en matchtechnologie. Het CBS heeft van Textkernel een bestand ontvangen met alle gewebscrapete vacatures van het 4e kwartaal van 2017 tot en met het 4e kwartaal van 2018. Het CBS heeft zowel de volledige vacatureteksten ontvangen als ook geclassificeerde gegevens. Textkernel classificeert de bedrijven naar de Standaardbedrijfsindeling (SBI) en de beroepen naar de International Standard Classification of Occupation (ISCO) en de Nederlandse vertaling daarvan (BRC).

Directe en intermediaire vacatures

Vacatures kunnen geplaatst worden door het bedrijf dat personeel vraagt (directe vacatures), maar ook door een intermediair, bijvoorbeeld een uitzendbureau. Bij een intermediaire (indirecte) vacature is het vaak niet duidelijk welk bedrijf personeel vraagt. Het is daarmee niet bekend in welke bedrijfstak personeel gevraagd wordt. Het beroep, de gevraagde competenties en de locatie staan wel in de vacaturetekst.

Dubbele vacatures verwijderd

Zoals al genoemd in de inleiding, worden vacatures meerdere keren op internet teruggevonden, omdat ze op verschillende websites tegelijk worden geplaatst. Soms worden vacatures ook meerdere keren op dezelfde website geplaatst als de vacature nog openstaat. Gemiddeld wordt een vacature 6,5 keer op internet teruggevonden door Textkernel. Al deze postings gaan over één vacature, en moeten daarom maar eenmaal meetellen in de statistiek. Duplicaten kunnen door het algoritme van Textkernel worden herkend door middel van analyse van de vacatureteksten.

Vervolgens worden de duplicaten verwijderd. Doordat Textkernel de duplicaten wel beschikbaar heeft gesteld aan het CBS, was het mogelijk om de kwaliteit van het ontdebellen (beknopt) te toetsen. De check gaf geen aanleiding tot twijfel aan de methode van ontdebelling.

In het algemeen geldt wel dat intermediaire vacatures lastiger te ontdebellen zijn dan directe vacatures. Een belangrijk kenmerk bij het onderscheiden van unieke vacatures is het bedrijf dat personeel zoekt, maar vaak ontbreekt identificerende informatie over het bedrijf bij intermediaire vacatures. Ook past de intermediair de vacaturetekst aan, waardoor de duplicaten niet altijd herkenbaar zijn. Soms wordt de vacature geplaatst door zowel het bedrijf dat personeel vraagt als door een of meerdere intermediairs. Ook dat maakt ontdebellen lastig. Dit is een bekend probleem bij aanbieders van gewebscrapete vacatures.

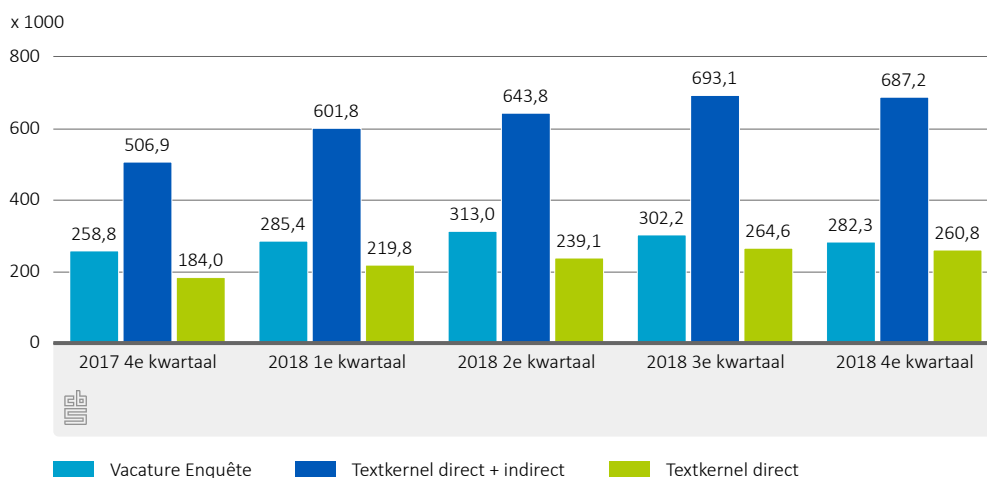
Aantal online vacatures sluit niet aan op Vacature Enquête

Ondanks de ontdebelling van vacatures sluit het aantal ontstane (nieuwe) vacatures vanuit webscraping niet aan op het aantal ontstane vacatures volgens de Vacature Enquête (VE) van het CBS. In het 4e kwartaal van 2017 is het aantal unieke vacatures van Textkernel met 507 duizend bijna tweemaal zo hoog als het aantal ontstane vacatures volgens de VE met 259 duizend. In de kwartalen daarna loopt het verschil verder op, zie Figuur 2.1.

Wanneer we alleen de directe vacatures van Textkernel selecteren, ligt het aantal ontstane vacatures zo'n 10–30 procent lager dan het aantal ontstane vacatures volgens de VE.

2.1 Ontstane vacatures Vacature Enquête en Textkernel

Ontstane vacatures



Bron: CBS, Textkernel

Online vacatures als aanvulling op VE

Uit bovenstaande vergelijking concluderen we dat de online vacatures niet gebruikt kunnen worden om de VE als bron te vervangen. Het niveau van de ontstane vacatures, maar ook de trend van online vacatures wijkt daarvoor teveel af van de VE. Hiervoor zijn een aantal redenen:

- Niet alle vacatures worden online gepubliceerd, zie publicatie Hoe werven werknemers (UWV, 2018);
- De algoritmes van Textkernel waarmee vacatures worden gezocht en ontdekt, veranderen (verbeteren) voortdurend;
- De technische specificaties van websites veranderen voortdurend waardoor vacatures plotseling juist wel of juist niet gevonden worden.

Deze ontwikkelingen zorgen ervoor dat er door de tijd heen meer of minder vacatures online door algoritmes gevonden kunnen worden. We houden daarom het aantal vacatures volgens de VE aan, maar willen de onderliggende informatie (over beroep, locatie, etc.) uit de online vacatures gebruiken om in meer detail over kenmerken van vacatures te kunnen publiceren.

2.2 Bewerking vacaturedata

Het doel van dit onderzoek is om vacatures per beroepsklasse te publiceren, waarbij het beroep afkomstig is uit online vacatures. We hebben een weegmodel nodig dat de online vacatures weegt, zodat het gewogen totaal online vacatures aansluit bij het totaal aantal ontstane vacatures volgens de VE. We kunnen simpelweg iedere online vacature hetzelfde gewicht geven, waarmee we ervan uit gaan dat in elke bedrijfstak en elk beroep een even groot aandeel vacatures online gepubliceerd wordt en even goed ontdekt wordt. Dit is echter niet waarschijnlijk.

In plaats daarvan wordt de informatie over de bedrijfstak uit de online vacature meegenomen om per bedrijfstak apart aan te sluiten bij de VE-cijfers. Daarnaast worden de VE-cijfers uitgebreid met vier beroepsniveaus, waardoor de online vacatures voor bedrijfstak én beroepsniveaus apart gewogen kunnen worden. De aanname van het weegmodel is vervolgens dat de

online gevonden vacatures binnen een kwartaal, bedrijfstak en beroepsniveau representatief zijn voor *alle* ontstane vacatures binnen hetzelfde kwartaal, bedrijfstak en beroepsniveau. De nodige aanpassingen op de vacaturedata worden in deze paragraaf beschreven.

2.2.1 Aanpassen bedrijfstak

Het CBS publiceert op StatLine de ontstane vacatures per bedrijfstak op basis van de VE. In de VE wordt aan werkgevers niet gevraagd in welke bedrijfstak ze actief zijn, maar de bedrijfstak wordt door het CBS toegekend aan de hand van het Algemeen Bedrijvenregister (ABR). Het ABR is een combinatie van registers, waarvan de belangrijkste het Nieuwe Handelsregister van de Kamer van Koophandel (KvK) betreft.

Ook Textkernel voegt een bedrijfstak toe aan de online vacatures. Hiervoor koppelt Textkernel het KvK-nummer en de SBI-code uit het Nieuwe Handelsregister aan de vacature, gebaseerd op de naam en locatie van het bedrijf volgens de vacaturetekst. Omdat de SBI zowel in de VE als in het Textkernelbestand bekend zijn, is mogelijk om de bedrijfstak op te nemen in het weegmodel, zodat iedere bedrijfstak apart aangesloten kan worden. Dit betekent dat een online vacature een hoger gewicht krijgt als deze hoort bij een bedrijfstak waar weinig online vacatures zijn gevonden, terwijl er volgens de VE wel veel vacatures zijn.

Het CBS heeft onderzocht in hoeverre de SBI-toekenning van Textkernel aan de vacatures correct is. Wanneer de SBI-toekenning niet correct is, kunnen daar drie redenen voor zijn:

1. Textkernel heeft het goede KvK-nummer gevonden bij het bedrijf uit de vacature, maar een andere SBI-code toegevoegd;
2. Textkernel heeft niet het juiste KvK-nummer toegevoegd aan het bedrijf en daardoor ook een verkeerde SBI-code toegekend;
3. Textkernel heeft op basis van de bedrijfsnaam en locatie geen KvK-koppeling kunnen maken, waardoor de SBI-code onbekend is.

Hieronder worden deze problemen nader bekeken.

SBI-code vervangen vanuit het ABR

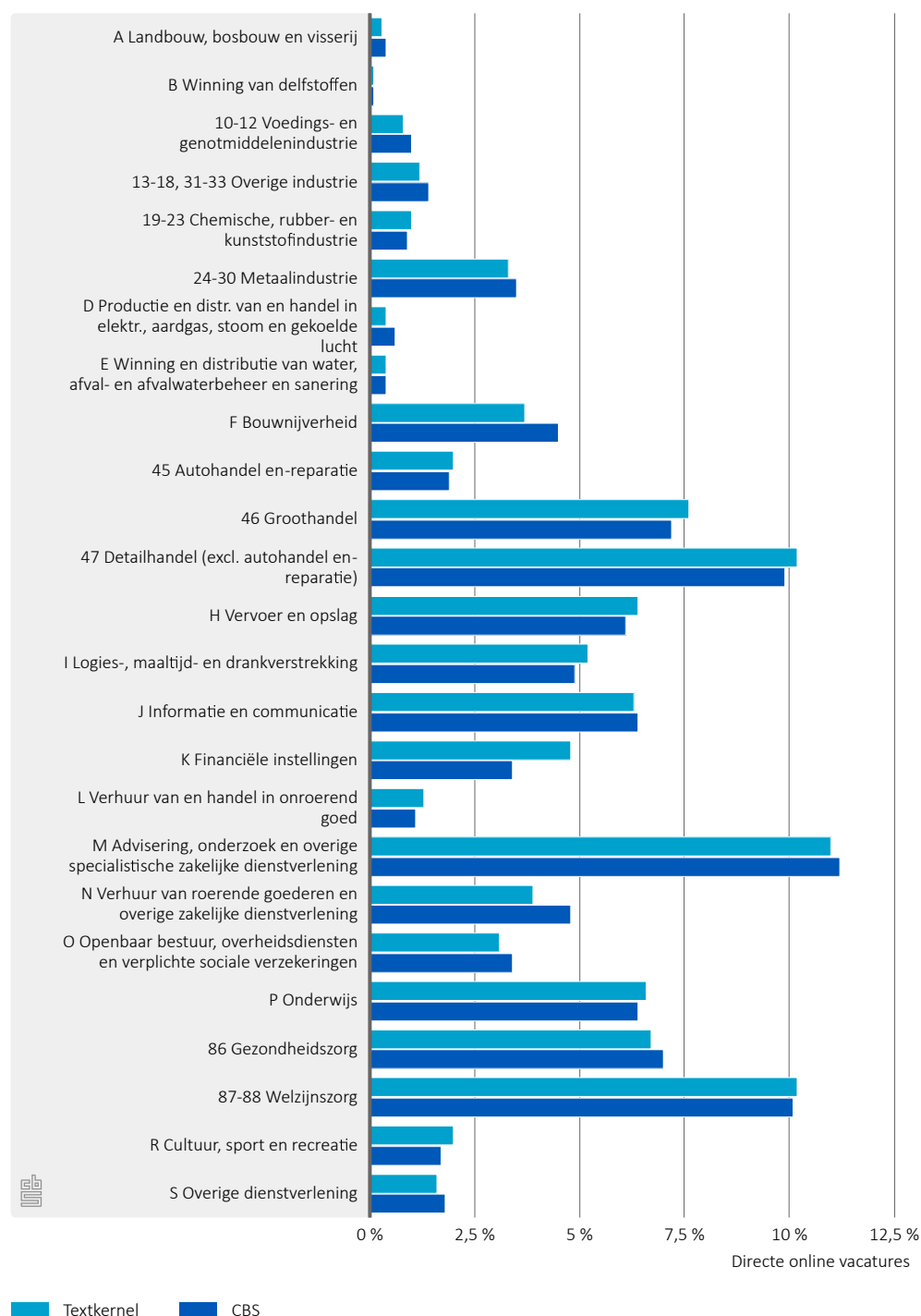
De toekenning van een SBI-code aan een juist KvK-nummer heeft het CBS onderzocht door de SBI-code uit het ABR te koppelen aan het KvK-nummer in het Textkernelbestand. Dit levert verschillen in de SBI-codes op, zoals te zien is in Figuur 2.2 voor de directe online vacatures.

Een van de redenen waarom de SBI-toekenning door Textkernel kan verschillen van de SBI uit het ABR van het CBS, is dat een bedrijf uit verschillende onderdelen kan bestaan met onderling verschillende SBI-codes. Textkernel heeft aan zo'n bedrijf één SBI-code toegekend, wellicht op basis van vacatures in het verleden, terwijl dit in de actuele situatie niet de grootste bedrijfseenheid is. Een voorbeeld is een middelgrote gemeente, die de SBI-toekenning 'preventieve gezondheidszorg' krijgt, in plaats van 'algemeen overheidsbestuur'. Ook lijkt het mogelijk dat de SBI-toekenningen van Textkernel gebaseerd zijn op KvK-informatie die inmiddels verouderd is.

Het ABR bevat de SBI van de hoofdactiviteit van een bedrijfseenheid. Wanneer een KvK-nummer in het bestand van Textkernel in het ABR uit meerdere bedrijfseenheden bestaat, hebben wij in dit onderzoek de vacatures met een kansmodel verdeeld over de verschillende bedrijfseenheden. Als een KvK-nummer bijvoorbeeld in het ABR uit twee bedrijfseenheden bestaat, namelijk

2.2 Toekenning bedrijfstak (SBI) door Textkernel en CBS bij directe online vacatures met bekende SBI, 2017Q4-2018Q4

Toekenning bedrijfstak



Bron: CBS, Textkernel

bedrijfseenheid A met 70 werknemers en B met 30 werknemers, dan wordt iedere vacature met een kans van 70 procent toebedeeld aan bedrijfseenheid A en met een kans van 30 procent aan bedrijfseenheid B. Op deze manier wordt rekening gehouden met bedrijven die met verschillende onderdelen in meerdere bedrijfstakken actief zijn. Overigens kan het CBS ook maar tot op zekere hoogte rekening houden met verschillende economische activiteiten binnen een bedrijf. Alleen de SBI-code van de hoofdactiviteit van een bedrijfseenheid wordt meegenomen, terwijl een bedrijfseenheid ook nevenactiviteiten in een andere bedrijfstak kan hebben.

De SBI-toewijzing aan bedrijven door Textkernel en CBS leidt op het eerste gezicht niet tot grote verschillen in aandeel vacatures per bedrijfstak op twee-digitniveau. Enkel bij de financiële instellingen (bedrijfstak K) heeft Textkernel een duidelijk groter aandeel vacatures dan het CBS. Bij de bouwnijverheid (bedrijfstak F) en verhuur van roerende goederen (bedrijfstak N) heeft Textkernel een wat kleiner aandeel vacatures dan het CBS. De overige verschillen op dit niveau zijn beperkt.

Wanneer we de bedrijfstakken in meer detail bekijken, worden de verschillen groter. Binnen dit onderzoek is de bedrijfstak Gezondheids- en welzijnszorg nader bekeken in het kader van het project Arbeidsmarkt Zorg en Welzijn (AZW). Er zijn ruim 6000 directe online vacatures die Textkernel indeelt in de bedrijfstak Gezondheids- en welzijnszorg, maar die het CBS indeelt buiten de bedrijfstak. Daarnaast zijn er ook ruim 7000 vacatures die Textkernel plaatst buiten Gezondheids- en welzijnszorg, maar die het CBS wel tot die bedrijfstak rekent. Netto geeft dat slechts een zeer klein effect in het aantal vacatures binnen de bedrijfstak Gezondheids- en welzijnszorg, maar dit zorgt wel voor verschuivingen in vacatures tussen de onderliggende bedrijfstakken.

Binnen die bedrijfstak zijn er wel SBI-verschuivingen met een groot effect op de aandelen vacatures binnen de verschillende zorg en welzijnsbranches, zie ook Figuur 2.3. Zo werden bijvoorbeeld alle vacatures van een universitair ziekenhuis (branche: UMC's) onterecht toegewezen aan een vestiging van kinderopvang van datzelfde ziekenhuis (branche: kinderopvang). Door de diverse verschuivingen tussen de branches zijn door het CBS onder andere het aantal vacatures van de UMC's meer dan verdubbeld en zijn de vacatures van het sociaal werk teruggebracht tot twee derde van het aantal van Textkernel.

Het CBS gebruikt voor het vervolg van dit onderzoek de SBI-code uit het ABR. Voor deze aanpak is gekozen omdat de SBI van Textkernel minder geschikt blijkt voor publicatie naar meer detail binnen bedrijfstakken. Wanneer het KvK-nummer uit het Textkernelbestand niet kan worden teruggevonden in het ABR (4% van de vacatures), wordt de SBI-code van Textkernel aangehouden.

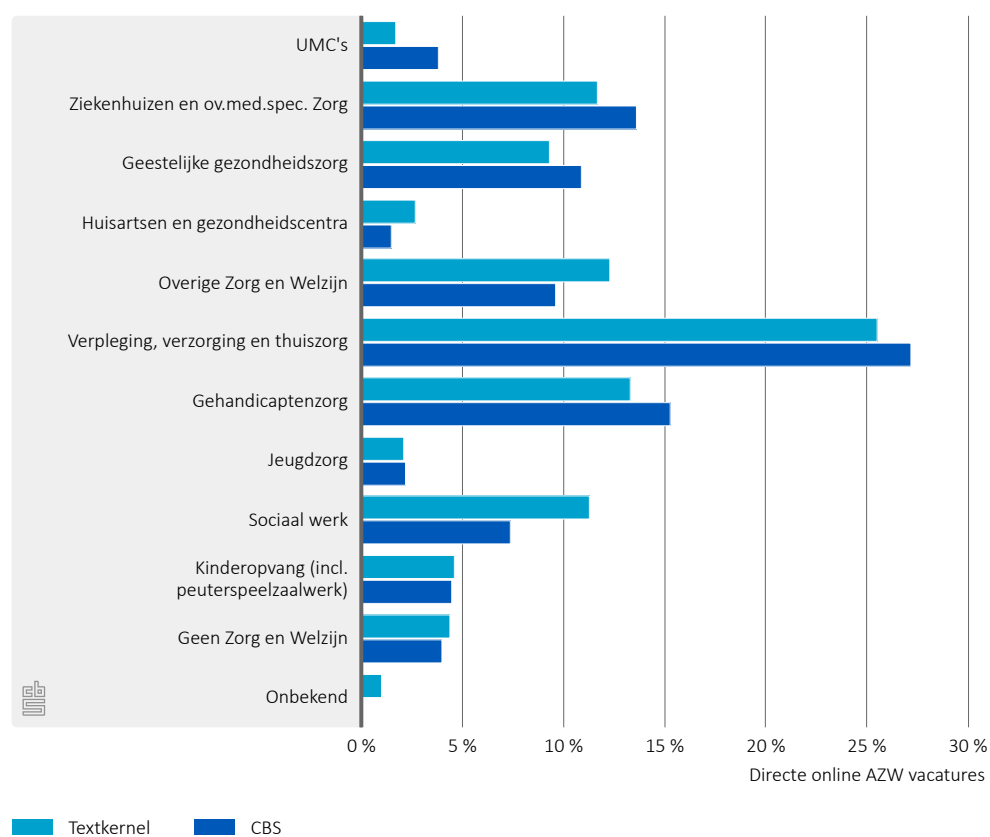
Handmatige correctie SBI na verkeerde KvK-koppeling

Een andere oorzaak waarom vacatures soms niet de goede SBI-code toegewezen gekregen hebben, is omdat Textkernel een onjuist KvK-nummer heeft gebruikt. Het is lastig om deze gevallen systematisch op te sporen en te corrigeren. In dit onderzoek hebben we naar de volgende twee indicatoren gekeken:

1. Onwaarschijnlijke verhouding van het aantal vacatures bij een bedrijf ten opzichte van het aantal werkzame personen volgens het ABR. Als een bedrijf bijvoorbeeld duizend vacatures heeft, terwijl er maar een aantal personen werken, dan is dit een indicatie dat het KvK-nummer niet correct is. De meest extreme gevallen hebben we binnen dit onderzoek onderzocht en handmatig aangepast.

2.3 Toekenning AZW hoofdbranche door Textkernel en CBS bij directe online AZW vacatures, 2017Q4-2018Q4

Toekenning AZW hoofdbranches



Bron: CBS, Textkernel

2. Beroepen komen in onwaarschijnlijke aantallen voor in de subbranches van een bedrijfstak. Er zijn bijvoorbeeld honderden vacatures van artsen en verpleegkundigen bij een culturele organisatie, doordat dit bedrijf het verkeerde KvK-nummer heeft gekregen. Binnen dit onderzoek hebben we alleen een aantal invloedrijke gevallen binnen de bedrijfstak Gezondheids- en welzijnszorg onderzocht en gecorrigeerd.

Handmatige toevoeging SBI bij onbekend KvK-nummer

Bij ruim 20 procent van de directe vacatures heeft Textkernel geen koppeling kunnen maken met het Nieuwe Handelsregister van de KvK, waardoor er geen SBI bekend is. Bij bedrijven met een groot aantal vacatures heeft het CBS handmatig via websites bepaald wat de SBI van het betreffende bedrijf is en deze toegevoegd aan het databestand.

2.2.2 Toevoegen beroepsniveau aan VE

Na bovenstaande exercitie kunnen we de bedrijfstak gebruiken in het weegmodel. In navolging van UWV willen we er echter niet zomaar vanuit gaan dat voor elk beroepsniveau binnen een bedrijfstak op dezelfde manier geworven wordt. Mogelijk worden vacatures met verschillende beroepsniveaus vaker of juist minder vaak online gepubliceerd, en deze niveaus moeten dan een lager of hoger gewicht krijgen in het weegmodel.¹⁾

Vanuit Textkernel is bekend wat het beroepsniveau van een vacature is. Textkernel bepaalt aan de hand van de functieomschrijving het beroep volgens de ISCO. Met de ISCO kunnen beroepen vervolgens ingedeeld worden in vier niveaus.

De VE bevat geen informatie over het beroep of het beroepsniveau van vacatures. Om toch per beroepsniveau apart te kunnen wegen, wordt de VE voor het weegmodel hiermee uitgebreid. Er wordt ervan uitgegaan, dat binnen iedere bedrijfstak de verhoudingen van de beroepsniveaus gelijk zijn bij het aantal werkenden en bij het aantal vacatures. De verhoudingen van de beroepsniveaus per bedrijfstak van de werkenden wordt verkregen uit de Enquête Beroepsbevolking (EBB). De VE-aantallen per bedrijfstak worden met deze verhoudingen onderverdeeld naar beroepsniveau, waardoor we ook de beroepsniveaus van de online vacatures kunnen gebruiken in het weegmodel.

Vooraf vacatures voor functies met een laag beroepsniveau krijgen hogere gewichten wanneer de beroepsniveaus worden toegevoegd aan de VE. Dat volgt uit een vergelijking tussen de resultaten van de weegmethode met en zonder deze toevoeging. Door de toevoeging stijgt het aandeel vacatures met het laagste beroepsniveau van 8% naar 12%, terwijl het aandeel vacatures met het hoogste beroepsniveau daalt van 32% naar 26%.

2.3 Weging

Om het aantal vacatures per beroep te kunnen bepalen, worden drie aparte weegfactoren berekend. Na het wegen van de online vacatures met deze weegfactoren, zal het aantal vacatures per bedrijfstak aansluiten op het aantal vacatures volgens de VE.

¹⁾ Uit de publicatie Hoe werven werknemers van (UWV, 2018) blijkt dat er verschillen zijn tussen de kanalen van vacatures waarvoor een hoog, middel of laag opleidingsniveau wordt vereist. Werkgevers werven laagopgeleide werknemers vaker lokaal en minder vaak via de eigen website.

1. Wegen voor vacatures van intermediairs, per kwartaal $\times$ BRC-Beroepssegment;
2. Wegen voor vacatures met onbekende bedrijfstak, per kwartaal $\times$ BRC-Beroepssegment;
3. Wegen om aan te sluiten op VE, per kwartaal $\times$ SBI-bedrijfstak $\times$ ISCO-Beroepsniveau.

De drie weegfactoren worden hieronder toegelicht.

2.3.1 Wegen voor vacatures van intermediairs

Zoals eerder aangegeven, wordt een deel van de vacatures direct door werkgevers geplaatst, en een ander deel door intermediairs. Bij intermediaire vacatures is het niet bekend in welke bedrijfstak het bedrijf actief is waar personeel gezocht wordt. Daarom kunnen de indirecte vacatures niet meegenomen worden in het weegmodel. Deze indirecte vacatures moeten echter niet gemist worden, omdat bijvoorbeeld een groot aantal werknemers met technische beroepen op deze manier worden geworven, zie ook (UWV, 2018).

Directe vacatures gewicht per beroepssegment

Om indirecte vacatures met een bekend beroep mee te nemen in de weging, krijgen de *directe* vacatures van het zelfde beroepssegment een zwaarder gewicht ter compensatie. Met andere woorden, directe vacatures wegen zwaarder wanneer deze bij een beroepssegment horen met veel indirecte vacatures. Maar omdat de bedrijfstak van de indirecte vacatures niet bekend is, moeten we de indirecte vacatures verdelen ongeacht de bedrijfstak. In deze eerste stap wordt ervan uitgegaan dat per beroepssegment de verhouding tussen directe en intermediaire vacatures gelijk is in verschillende bedrijfstakken.

We kiezen ervoor om niet volledig te compenseren voor alle indirecte vacatures. De indirecte vacatures bevatten meer duplicaten, ook na ontdubbeling (zie Paragraaf 2.1). Deze extra advertenties voor dezelfde vacature willen we niet meetellen. Er is in dit onderzoek ervoor gekozen om intermediaire vacatures voor 25 procent mee te tellen bij de directe vacatures van hetzelfde beroepssegment, een keuze die hieronder toegelicht wordt.

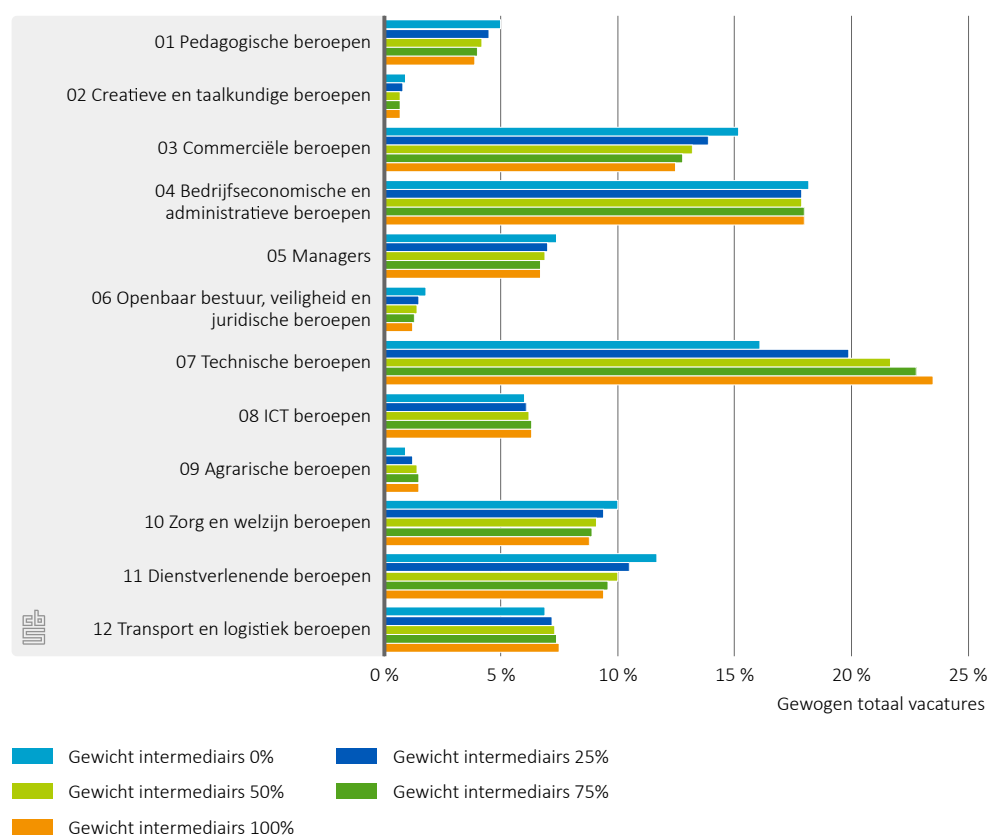
Intermediaire vacatures tellen voor 25 procent mee

Om de keuze van een gewicht voor intermediaire vacatures te kunnen beoordelen, zijn verschillende varianten toegepast, waarbij de intermediaire vacatures meetellen voor 0 procent (niet meetellen), 25, 50, 75 of 100 procent (net zo zwaar als een directe vacature). Figuur 2.4 laat de invloed zien van het gewicht van intermediaire vacatures op de verdeling van vacatures over beroepsklassen. De beroepsklassen met veel stijging hebben veel intermediaire vacatures, waaronder met name de beroepsklasse Technische beroepen. Wanneer intermediaire vacatures een gewicht van 0 procent krijgen, leidt het weegmodel tot een aandeel van 16 procent bij Technische beroepen op het totaal aantal vacatures, vergeleken met 24 procent wanneer indirecte vacatures voor 100 procent meegeteld worden. Wanneer intermediaire vacatures voor 25 procent meetellen, is het aandeel Technische beroepen 20 procent van de vacatures, rond het gemiddelde van de twee extremen.

Deze laatste observatie is ook een patroon in andere beroepsklassen. Voor vijf klassen (03, 06, 07, 09, en 11) geldt dat de helft van het totale effect van het gewicht intermediairs bereikt wordt bij een gewicht tussen 20 en 25 procent. Voor vier andere klassen (02, 05, 10, en 12) geldt dat de helft van het effect bereikt wordt bij een gewicht tussen 25 en 30 procent. Alleen bij de Pedagogische beroepen en de ICT-beroepen ligt het bij een gewicht iets boven de 30 procent.

2.4 Invloed gewicht intermediaire vacatures op verdeling vacatures per BRC-Beroepsklasse, 2017Q4-2018Q4

Gewicht intermediaire vacatures en beroepsklassen



Bron: CBS, Textkernel

De beroepsklasse Bedrijfseconomische en administratieve beroepen is een uitzondering. Het aandeel van deze beroepsklasse is bij alle varianten 18 procent.

De observaties onderbouwen de keuze van een gewicht voor intermediaire vacatures van 25 procent, als gemiddelde uitkomst tussen het volledig meetellen en het volledig negeren van intermediaire vacatures.

2.3.2 Wegen voor vacatures met onbekende bedrijfstak

In de tweede stap wordt een weegfactor berekend om directe vacatures met een onbekende SBI-code op te tellen bij de directe vacatures met bekende SBI-code, opnieuw apart voor ieder BRC-Beroepssegment.

2.3.3 Wegen voor aansluiting VE

In een laatste stap wordt een derde weegfactor berekend door het gewogen totaal aantal vacatures met bekende SBI-code te laten aansluiten op het totaal aantal ontstane vacatures volgens de VE, per SBI-bedrijfstak en per BRC-Beroepsniveau.

3 Rol intermediaire vacatures in het weegmodel

UWV maakt al vanaf 2012 gebruik van online vacatures van Textkernel om arbeidsmarktanalyses uit te voeren. Toen het CBS startte met dit onderzoek, had UWV samen met Panteia al een methode ontwikkeld om de online vacatures te laten aansluiten op uitkomsten van de VE, zoals die gepubliceerd worden op StatLine. Het CBS heeft voortgeborduurd op de bevindingen van UWV, en heeft de methode op een aantal punten gewijzigd, zodat de aansluiting met de VE verder wordt verbeterd.

Het resultaat van dit onderzoek is een weegmodel met in de basis een gelijke opzet als die van UWV, maar met een aantal wijzigingen op de details van de methode. De wijzigingen komen deels door verschillen in de beschikbaarheid van data en aangepaste onderzoeksdoeleinden, hierop wordt in Paragraaf 3.1 ingegaan.

UWV berekent ook weegfactoren, wel in een andere volgorde, waardoor intermediaire vacatures later worden meegeteld. Dit verschil heeft invloed op de uitkomsten van het model, wat betreft de verdeling van beroepen en de verdeling van bedrijfstakken. Het gevolg van het verschil is dat de totalen bij methode UWV niet meer aansluiten op de VE-totalen van de bedrijfstakken. In Paragraaf 3.2 wordt de invloed van de volgorde bij de berekening van de weegfactoren verder toegelicht.

Het hoofdstuk eindigt met een conclusie over de gekozen methode van het CBS.

3.1 Verschillen weegmethodiek UWV en CBS

UWV gebruikt al jaren de online vacatures van Textkernel om meer inzicht te krijgen in ontwikkelingen op de arbeidsmarkt. Het CBS sluit daar in deze publicatie grotendeels bij aan. Toch zullen de uitkomsten op basis van het CBS-onderzoek afwijken van eerder door het UWV gepubliceerde cijfers. Allereerst heeft dit te maken met het moment waarop beide organisaties het onderzoek uitvoeren: direct na afloop van een kwartaal of na afloop van een jaar. Daarnaast is de mate van detail waarop gepubliceerd wordt van invloed. Het UWV publiceert op heel gedetailleerd niveau over vacatures per beroep. Voor het CBS is de combinatie van beroep en bedrijfstak belangrijk.

De laatste twee verschillen hieronder betreffen voorstellen van het CBS om de methode te verbeteren.

Andere versie vacaturedata

UWV heeft van Textkernel een ander bestand met online vacatures ontvangen dan het CBS. UWV krijgt na afloop van een kwartaal een bestand met de gevonden vacatures in dat kwartaal. Het CBS heeft begin 2019 eenmalig een bestand ontvangen van Textkernel met alle gevonden vacatures van het 4e kwartaal van 2017 tot en met het 4e kwartaal van 2018. Wanneer Textkernel wijzigingen doorvoert in de techniek van webscraping en ontdebelling, wordt dit met terugwerkende kracht gedaan. Ook informatie over oude kwartalen verandert daardoor. Het CBS heeft dus over dezelfde periode een bestand met een ander aantal vacatures ontvangen dan UWV.

Voortschrijdend gemiddelde versus gemiddelde van vijf kwartalen

Het CBS streeft naar één aanvullende statistiek per jaar, terwijl UWV per kwartaal publiceert. Om de resultaten robuust te maken berekent UWV weegfactoren op basis van voortschrijdende gemiddeldes, waarbij de vacatures van voorgaande kwartalen met een kleiner gewicht meetellen.

BRC Beroepsgroep versus Beroepssegment

UWV corrigeert voor de intermediaire vacatures apart voor ieder BRC Beroepsgroep, terwijl het CBS corrigeert op het hoger niveau van het BRC Beroepssegment. Dit komt omdat UWV als doel heeft om te kunnen publiceren per beroep, terwijl het CBS doelt op meer robuustheid per beroepssegment wanneer het gaat inzoomen op details per bedrijfstak.

CBS gebruikt SBI-code vanuit ABR

UWV gebruikt de SBI-toewijzing van Textkernel, terwijl het CBS ervoor heeft gekozen om eigen bedrijfsinformatie uit het ABR over de bedrijfstak te verwerken op basis van het KvK-nummer (zie Paragraaf 2.2.1).

CBS voegt intermediaire vacatures eerder toe in weegmodel

UWV voert dezelfde stappen in het weegmodel uit als het CBS (zie Paragraaf 2.3), maar in een andere volgorde. UWV weegt eerst om de directe online vacatures aan te laten sluiten op de VE, en voegt daarna nog intermediaire vacatures toe. In een extra stap wordt het totaal per kwartaal weer teruggebracht naar het VE-totaal, maar het gevolg is dat bij UWV het gewogen

aantal vacatures *per bedrijfstak* niet meer aansluit op de VE. Het CBS voegt echter eerst een weegfactor toe waarmee intermediaire vacatures meegewogen worden, en laat daarna de directe en indirecte vacatures samen aansluiten op de VE. In de volgende paragraaf wordt dit verschil in volgorde nader toegelicht.

3.2 Invloed aanpassing volgorde weegstappen

3.1 Verschil volgorde weegstappen CBS en UWV

Weegstappen	Weegvolgorde	
	CBS	UWV
Wegen voor vacatures van intermediairs, per kwartaal × BRC-Beroepssegment	Stap 1	Stap 3
Wegen voor vacatures met onbekende bedrijfstak, per kwartaal × BRC-Beroepssegment	Stap 2	Stap 1
Wegen om aan te sluiten op VE, per kwartaal × SBI-bedrijfstak × ISCO-Beroepsniveau	Stap 3	Stap 2
Wegen om aan te sluiten op VE, per kwartaal	n.v.t.	Stap 4

Welke invloed heeft het aanpassen van de volgorde van de weegstappen op de resultaten? Om dit te onderzoeken voeren we de weegstappen in twee verschillende volgorden uit, waarbij we overige verschillen tussen de weegmethodiek van het CBS en UWV buiten beschouwing laten. De twee weegvolgordes worden dus op dezelfde dataset uitgevoerd, met de SBI-toewijzing vanuit het ABR. Intermediaire vacatures worden in beide gevallen voor 25 procent meegeteld. De enige wijziging is de stap waarin de intermediaire vacatures toegevoegd worden.

Op die manier kan worden onderzocht, welk invloed alleen de volgorde van de weegstappen heeft op de resultaten per bedrijfstak en de resultaten per beroepsklasse. De twee te vergelijken volgordes van de weegstappen zijn nog eens weergegeven in Tabel 3.1.

3.2.1 Bedrijfstakken

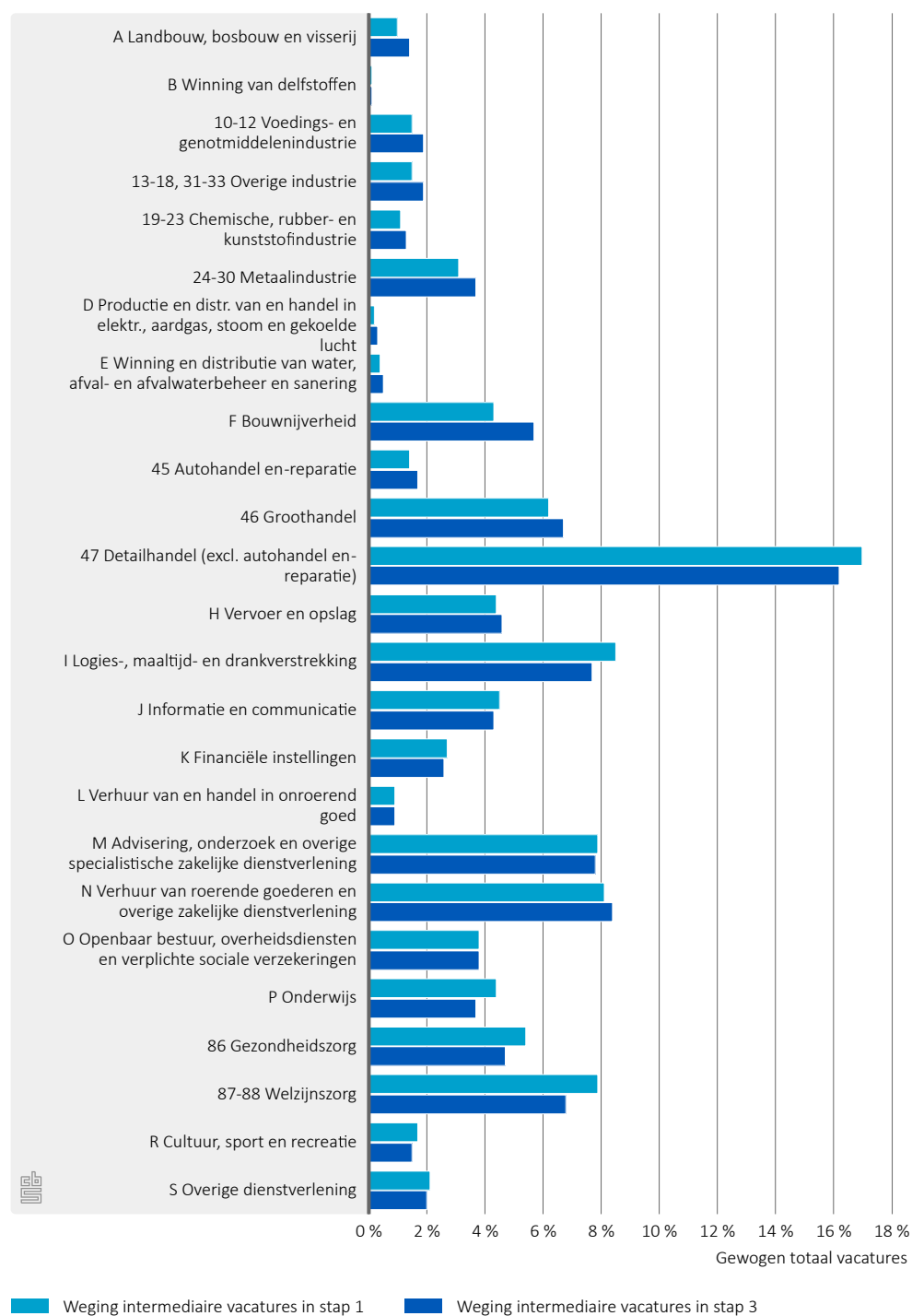
Wanneer eerst gewogen wordt voor vacatures van intermediairs, sluiten uiteindelijk de vacatures per bedrijfstak aan op het aantal vacatures volgens de VE. Wanneer pas in stap 3 gewogen wordt voor intermediairs, sluit alleen het totaal aantal vacatures aan op de VE, maar per bedrijfstak wijken de aantallen fors af zoals te zien in Figuur 3.1.

Later toevoegen van intermediaire vacatures zorgt voor stijging vacatures Bouwnijverheid, daling Onderwijs

Wanneer pas in stap 3 gewogen wordt voor intermediaire vacatures, stijgt het aantal vacatures in de bedrijfstak Bouwnijverheid met ruim 30 procent ten opzichte van de VE. Omdat het totaal aantal vacatures blijft aansluiten, wordt deze stijging gecompenseerd met een daling in andere bedrijfstakken. In de bedrijfstak Onderwijs daalt het aantal vacatures daarom met 15 procent ten opzichte van de VE.

3.1 Invloed volgorde weegstappen op verdeling vacatures per bedrijfstak (SBI), 2017Q4-2018Q4

Bedrijfstakken bij weegvolgordes



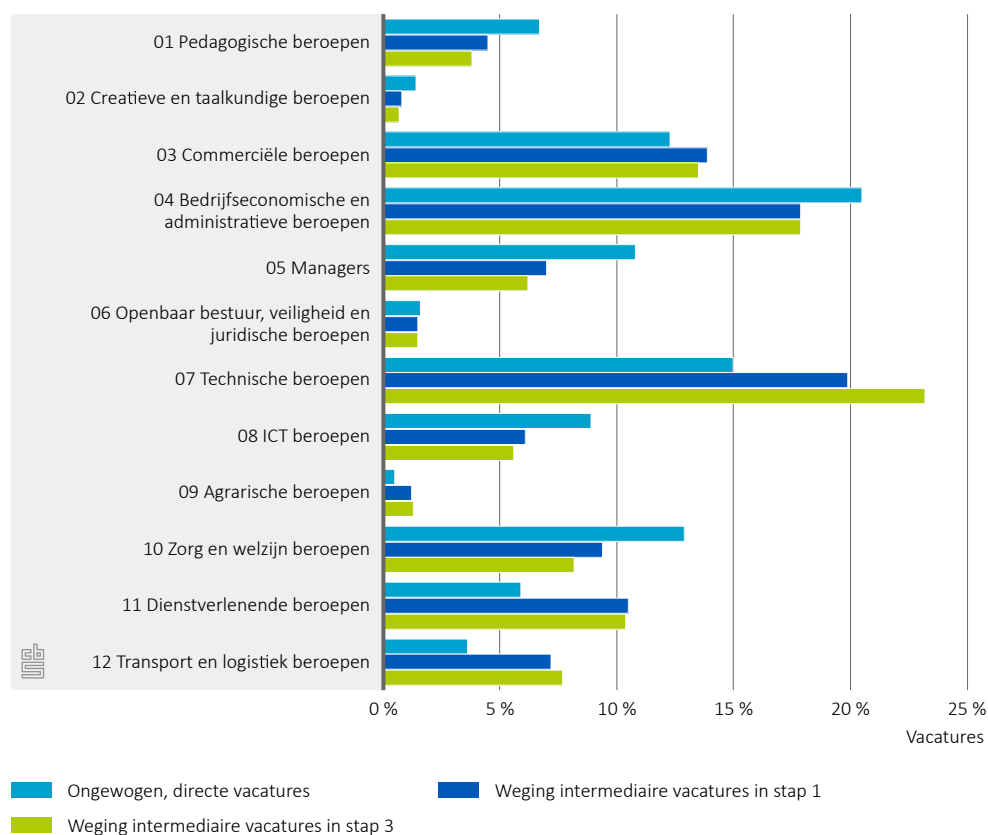
Bron: CBS, Textkernel

3.2.2 Beroepsklassen

De volgorde van de weegstappen heeft ook een grote invloed op de vacatures per beroepsklasse. Figuur 3.2 laat de verdeling van vacatures naar beroepsklassen zien. De directe vacatures bevatten (ongewogen) relatief veel vacatures van bedrijfseconomische en administratieve beroepen en managers en relatief weinig technische en dienstverlenende beroepen. De intermediaire vacatures bevatten juist relatief veel technische en dienstverlenende beroepen.

3.2 Invloed weging op verdeling vacatures per BRC-Beroepsklasse, 2017Q4-2018Q4

Beroepsklassen bij weging



Bron: CBS, Textkernel

Later toevoegen leidt tot grotere invloed intermediaire vacatures

Toevoeging van de intermediaire vacatures in stap 3 leidt tot een sterkere ophoging van de technische beroepen, en tot een grotere daling van andere beroepsklassen, zoals pedagogische beroepen, managers en zorg en welzijn beroepen.

Een mogelijke verklaring van de grotere extremen bij het wegen voor intermediairs in stap 3, is dat dubbelop wordt gecorrigeerd voor indirecte vacatures. Bij deze weegvolgorde wordt in stap 2 al aangesloten op de VE. De achterliggende gedachte bij het aansluiten op de VE is dat niet alle vacatures online waargenomen worden, en dat daar dus voor moet worden gecorrigeerd. Wanneer een bedrijfstak echter voor een groot deel bestaat uit beroepen waarvoor indirect wordt geadverteerd via intermediairs, dan is het aantal directe vacatures klein in verhouding. In dit geval krijgen de directe vacatures in stap 2 een groot gewicht om aan te sluiten bij de VE. Eigenlijk zijn de indirecte vacatures dan deels al verwerkt in deze tweede stap.

3.2 Aandeel vacatures en gemiddelde weegfactoren van de top 90 procent meest voorkomende BRC-Beroepssegmenten binnen bedrijfstak Bouwnijverheid, 2017Q4-2018Q4

		Aandeel vacatures		Weegfactor	
		Weging intermediairs in	Weging intermediairs in	Weging intermediairs in	Weging intermediairs in
Beroepssegment		stap 1	stap 3	stap 1	stap 3
073	Bouwarbeiders	20%	22%	3,3	4,6
076	Elektriciens en elektronici	17%	18%	2,5	3,6
074	Metaalarbeiders, machinisten	16%	17%	2,3	3,2
072	Vakspecialisten natuur en techniek	11%	10%	0,9	1,1
043	Administratief personeel	6%	6%	1,6	2,2
121	Bestuurders voertuigen en bedieners mobiele machines	4%	4%	2,7	3,8
078	Hulpkrachten bouw en industrie	4%	5%	5,9	10,3
071	Ingenieurs en onderzoekers wis-, natuur- en technische wetenschappen	4%	3%	0,9	0,9
053	Managers productie en gespecialiseerde dienstverlening	3%	2%	0,8	0,8
081	Specialisten ICT	2%	2%	0,9	0,9
041	Specialisten bedrijfsbeheer en administratie	2%	1%	0,8	0,7
033	Verkopers	1%	2%	1,4	1,9

Wanneer echter al in de eerste stap gecorrigeerd wordt voor vacatures van intermediairs, wordt bij het aansluiten op de VE alleen maar nog gecorrigeerd voor vacature die helemaal niet online verschenen of gevonden zijn.

Weinig verschil beroepsverdeling binnen bedrijfstakken

Wat is het effect van het vasthouden van de totalen uit de VE per bedrijfstak op de beroepsverdeling binnen bedrijfstakken? We onderzoeken dit voor de bedrijfstak Bouwnijverheid, waar de invloed van de weegvolgorde het grootst is. Toch laat Tabel 3.2 zien dat de aandelen van de belangrijkste beroepen binnen de bedrijfstak Bouwnijverheid nauwelijks verschillen tussen beide weegvolgordes.

De weegfactoren in de bedrijfstak Bouwnijverheid zijn uiteraard lager wanneer de intermediairs al in stap 1 toegevoegd worden in plaats van in stap 3. In Paragraaf 3.2.1 was al te zien dat het wegen voor intermediairs in stap 3 leidt tot 30 procent meer vacatures in de Bouwnijverheid. Het wegen voor intermediairs in stap 1 leidt echter niet vaker tot een weegfactor kleiner dan 1, met uitzondering van het beroepssegment Vakspecialisten natuur en techniek.

3.3 Conclusie methode CBS

Online vacatureteksten bevatten een grote schat aan informatie over de functie en het bedrijf dat de vacature uitzet. Door middel van tekstherkenningssoftware kan bijvoorbeeld het gevraagde beroep, het opleidingsniveau en de werklocatie herkend en geclassificeerd worden. Omdat de online vacatures echter geen compleet beeld van alle vacatures op de Nederlandse arbeidsmarkt schetsen, heeft het CBS een methode ontwikkeld om de online vacatures te wegen naar het aantal vacatures volgens de Vacature Enquête van het CBS.

Het CBS heeft een weegmodel ontwikkeld dat online vacature data combineert met data uit enquêtes en registers. Het doel van het model is om een totaalbeeld te kunnen schetsen over gevraagde beroepen in vacatures. Omdat de online vacatures geen compleet geven van de vacatures op de Nederlandse arbeidsmarkt, worden voor het totaal aantal vacatures per branche de aantallen vanuit de VE aangehouden. Dit wordt bereikt in de methode van het CBS door alle corrigerende stappen, inclusief de correctie voor intermediaire vacatures, uit te voeren voor dat in een laatste stap wordt aangesloten op de VE-totalen per bedrijfstak.

Methode UWV doorontwikkeld

Bij de ontwikkeling van het weegmodel kon het CBS voortborduren op de methode van UWV. De opbouw van het model is overgenomen van UWV, en ook in detail zijn er veel overeenkomsten zoals het uitbreiden van de VE-totalen met beroepsniveaus. Een cruciaal verschil is dat UWV eerst aansluit bij de VE en dan pas corrigeert voor intermediaire vacatures, waardoor de totalen per bedrijfstak afwijken van de VE. Hierdoor verandert het totaalbeeld vacatures per beroep, ook al zijn de verhoudingen binnen een bedrijfstak robuust.

Ook vervangt het CBS bedrijfstakinformatie uit het Textkernel bestand met eigen registerdata. Dit is nodig omdat de toewijzing van de bedrijfstak vooral in detail regelmatig verschilt. Door eigen registerdata te gebruiken wordt aangesloten op de bedrijfsindeling van de VE. Daarnaast zijn bij de opvallendste onwaarschijnlijke uitkomsten de KvK-nummers handmatig gecheckt en indien nodig gecorrigeerd. Dit leidt tot robuuste cijfers op niveau van bedrijfstak en beroeps-klasse.

Vervolgonderzoek

Er zijn een aantal manieren waarop het weegmodel verder verbeterd of uitgebreid zou kunnen worden. Een aantal suggesties voor vervolgonderzoek zijn:

- bedrijfstak intermediaire vacatures toevoegen. Omdat bij intermediaire vacatures het bedrijf dat personeel vraagt, onbekend is, is ook de bedrijfstak onbekend. Wellicht is het mogelijk om de bedrijfstak via textmining af te leiden uit de vacatureteksten.
- bulkvacatures onderscheiden. Soms wordt één vacatureposting geplaatst door een werkgever (of intermediair) voor meerdere vacatures met dezelfde functie. Denk bijvoorbeeld aan een advertentie van een restaurant voor een medewerker in de bediening, waar meerdere personen voor gezocht worden. Is dit uit de vacaturetekst af te leiden?
- het weegmodel kan uitgebreid worden met andere kenmerken van vacatures, zoals de werklocatie of de gevraagde competenties.

In dit onderzoek zijn de vacatures voor de bedrijfstak Zorg en Welzijn verder geanalyseerd. In deze bedrijfstak heeft het CBS meer correcties uitgevoerd, waarmee het mogelijk is om in meer detail over dit bedrijfstak te publiceren. Bij gedetailleerde publicatie van andere bedrijfstakken zal de plausibiliteit van de resultaten opnieuw onderzocht moeten worden en zullen er waarschijnlijk verdere handmatige correcties nodig zijn.

4 Resultaten

In dit hoofdstuk worden de eerste uitkomsten getoond op basis van het weegmodel dat beschreven is in Hoofdstuk 2. Het primaire doel van dit onderzoek was om het aantal ontstane vacatures per beroepsklasse te berekenen. De resultaten hiervan worden in Paragraaf 4.1 beschreven. Er is echter vraag naar meer detail. Daarom worden in Paragraaf 4.2 ingezoomd op de bedrijfstak Zorg en Welzijn, op verzoek van het ministerie van Volksgezondheid, Welzijn en Sport (VWS).

De resultaten in in dit hoofdstuk zijn gebaseerd op online vacatures van Textkernel, herwogen naar het aantal vacatures volgens de Vacature Enquête (VE). De SBI-code is zoveel mogelijk afkomstig van het CBS, er zijn handmatige correcties uitgevoerd op de KvK-nummers (zie Paragraaf 2.2.1) en de intermediaire vacatures zijn voor 25 procent meegeteld (zie Paragraaf 2.3.1). Er is eerst gewogen voor intermediaire vacatures, en daarna pas aangesloten op de VE. De uitkomsten wijken daarom af van eerder gepubliceerde vacatures per beroep van UWV (zie Hoofdstuk 3).

4.1 Beroepsklassen

In de periode 2017Q4 tot en met 2018Q4 ontstonden 1,44 miljoen nieuwe vacatures. Het aantal ontstane vacatures per kwartaal varieerde van 259 duizend in 2017Q4 tot 313 duizend vacatures in 2018Q2 (zie Figuur 2.1).

In ieder kwartaal was de vraag naar technische beroepen het grootst en de vraag naar creatieve en taalkundige beroepen het kleinste, zoals te zien in Figuur 4.1. De verhoudingen tussen de beroepsklassen is vrij stabiel over de tijd. Meerdere beroepsklassen volgen een algemeen seizoenspatroon: het aantal vacatures neemt toe, met een maximum in de lente, en minder tegen het einde van het jaar.

Bij dienstverlenende beroepen is het aantal ontstane vacatures duidelijk hoger in de lente en zomer. Dit wordt onder meer veroorzaakt door een piek in horecavacatures in het zomerseizoen. De vraag naar pedagogische beroepen is met name hoger in het tweede kwartaal, voor aanvang van het nieuwe schooljaar.

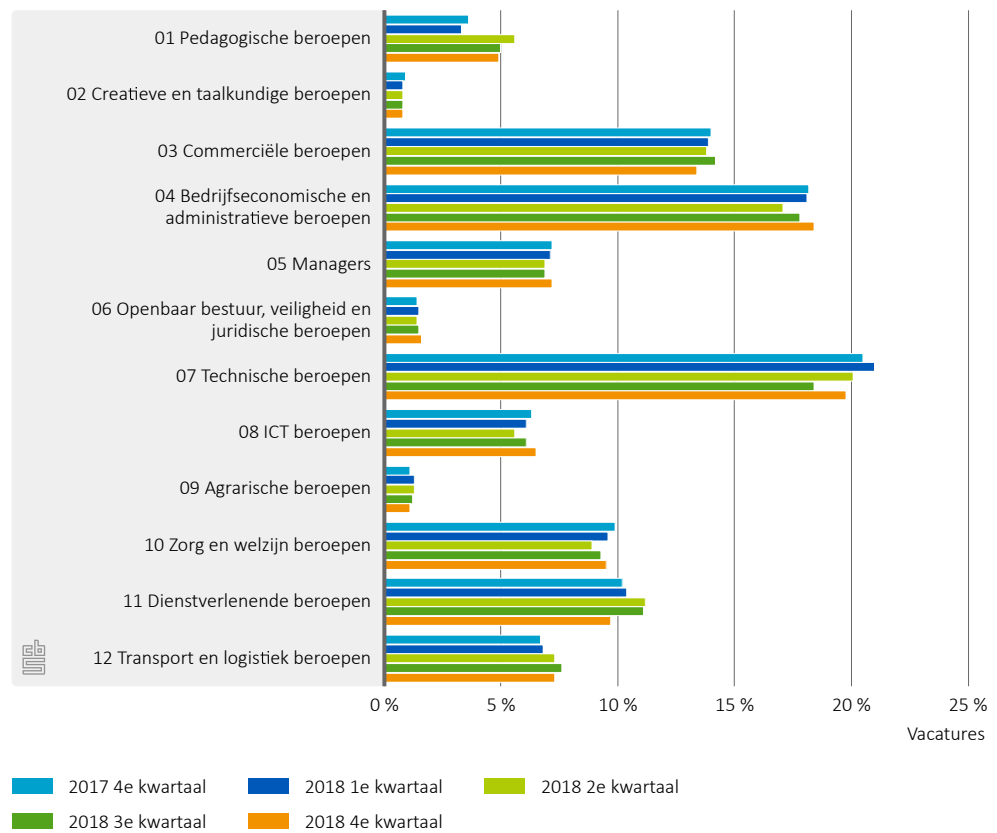
4.2 Vacatures Zorg & welzijn naar branche en naar regio

Uit de VE van het CBS blijkt dat er 191 duizend vacatures zijn ontstaan in de bedrijfstak Zorg en Welzijn in de vijf onderzochte kwartalen (4e kwartaal 2017 tot en met 4e kwartaal 2018). Figuur 4.2 laat zien dat het aantal ontstane vacatures voor zorg en welzijn oploopt van 36 duizend vacatures in 2017Q4 tot 40 duizend in 2018Q3.

Via webscraping worden er in deze periode 155 duizend vacatures voor zorg en welzijn online gevonden, dat is ruim 81 procent. Voor aansluiting bij de VE wordt het totaal van de online vacatures bijgesteld. Een online vacature in de zorg telt dan gemiddeld voor 1,2 vacatures, waarbij wel onderscheid wordt gemaakt tussen beroepsniveau van de vacatures.

4.1 Verdeling vacatures per BRC-Beroepsklasse en kwartaal, 2017Q4-2018Q4

Vacatures per beroepsklasse

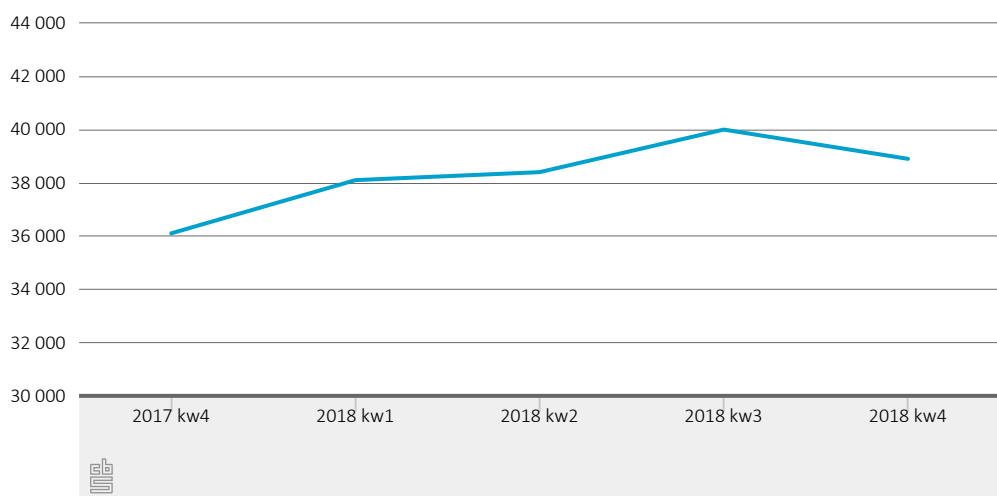


Bron: CBS, Textkernel

4.2 Vacatures in Zorg en Welzijn

Vacatures Zorg en Welzijn (incl kinderopvang)

Ontstane vacatures (kwartaal)

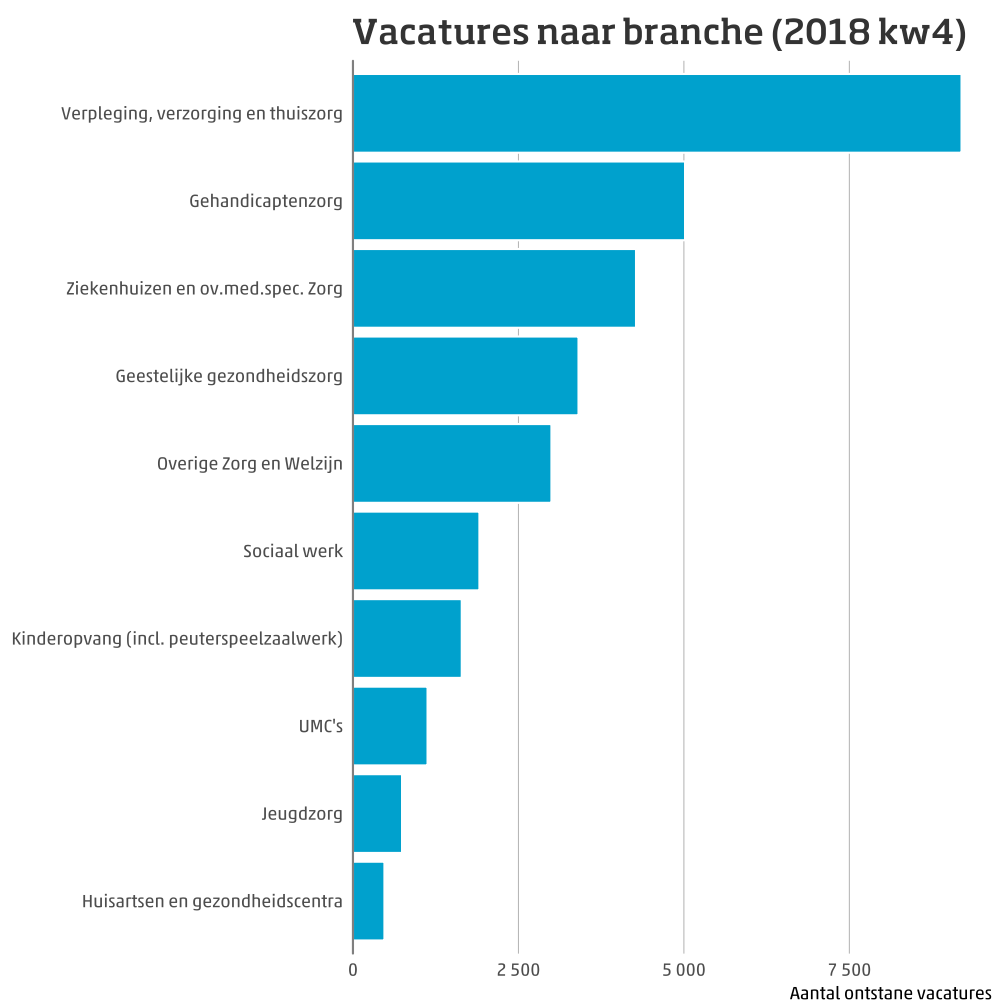


Uit dit onderzoek komen resultaten die ook naar kleinere groepen binnen de zorg te publiceren zijn, omdat er veel meer individuele vacatures direct worden waargenomen dan bij de VE. Zo hebben we voor de bedrijfstak Zorg en Welzijn de vacatures naar hoofdbranches uitgesplitst en hebben we de vacatures op basis van postcodegegevens ook toegedeeld aan RegioPlus-arbeidsmarktregio's.

4.2.1 Vacatures Zorg en Welzijn naar AZW-hoofdbranches

De meeste vacatures komen vanuit de branche verpleging, verzorging en thuiszorg, zie Figuur 4.3. In het laatste kwartaal van 2018 zijn daar ruim 10 duizend vacatures verschenen. Dat komt ook overeen met de verwachting: de branche verpleging, verzorging en thuiszorg heeft de meeste werknemers binnen de zorg en welzijn. De minste vacatures komen vanuit de Huisartsen en gezondheidscentra. Dit zijn over het algemeen kleine instellingen met weinig werknemers. De praktijkondersteuner van de huisarts heeft waarschijnlijk wel een werknemersaanstelling, een huisarts in een maatschap is geen werknemer.

4.3 Verdeling vacatures Zorg en Welzijn per AZW hoofdbranche



4.2.2 Vacatures Zorg en Welzijn naar RegioPlus-arbeidsmarktregio's

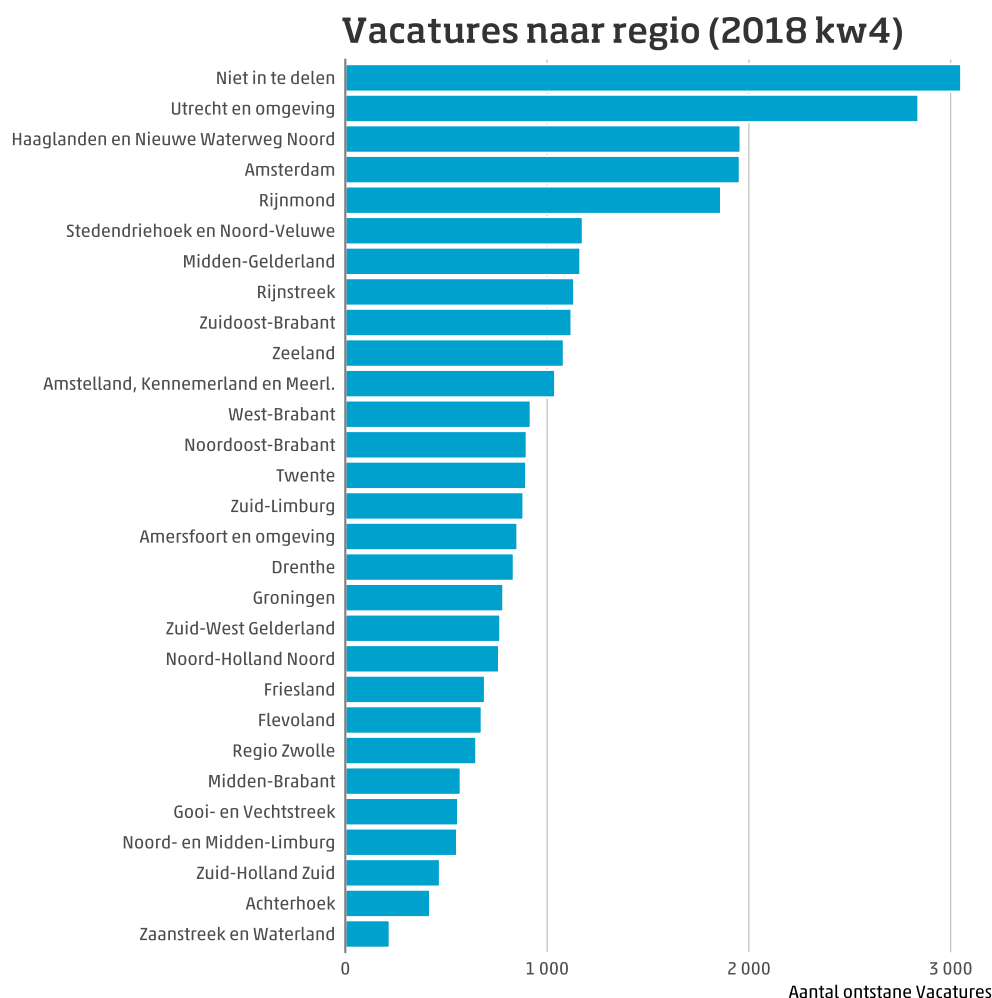
Uit de standplaats van de vacature is vaak af te leiden in welke regio gezocht wordt. Dit kan in 90% van de vacatures in de zorg en welzijn afgeleid worden, er blijft een groep over van 18 duizend vacatures zonder vermelde standplaats en die daardoor niet aan een regio toegedeeld

kan worden. De vacatures die aan regio's toegedeeld kunnen worden zijn in Figuren 4.4 en 4.5 in kaart gebracht.

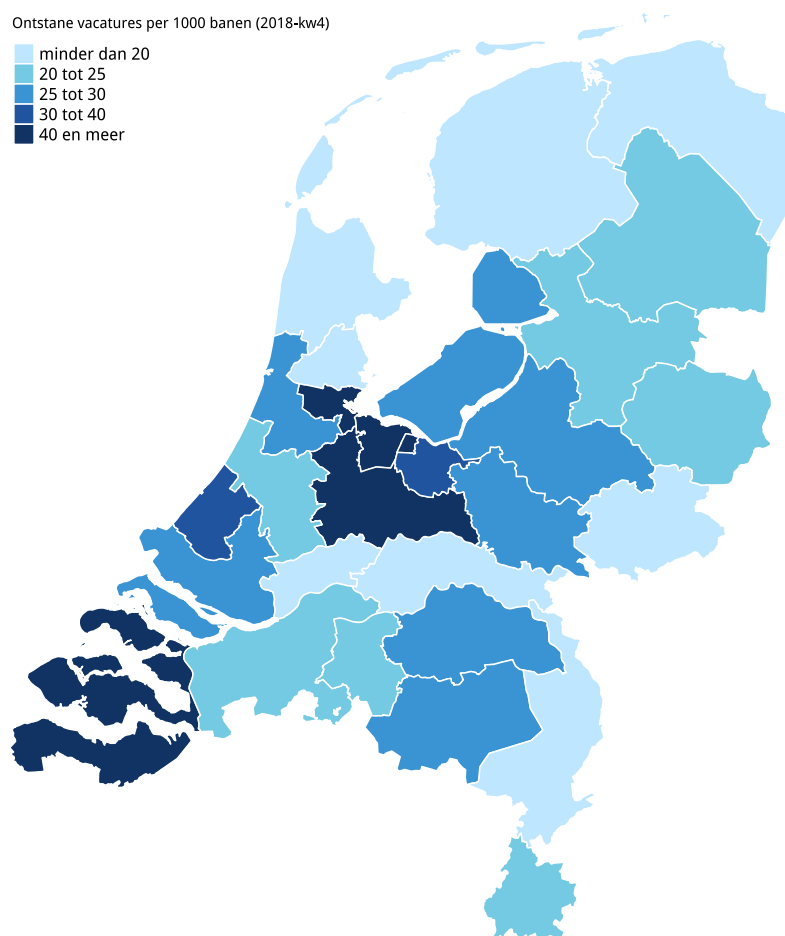
De meeste vacatures komen uit de RegioPlus-arbeidsmarktregio Utrecht en omgeving. In het vierde kwartaal van 2018 zijn er daar 3,7 duizend verschenen.

RegioPlus-arbeidsmarktregio's verschillen van grootte: in Rijnmond zijn er 86 duizend banen in de zorg en welzijn (2018, kwartaal 4), in Gooi en Vechtstreek zijn er 18 duizend banen. De regio's vergelijken we daarom in verhouding tot het aantal banen. De meeste vacatures in vergelijking met het aantal banen zijn er in Amsterdam met 46,6 ontstane vacatures per 1000 banen (2018, kwartaal 4). Opvallend laag is de regio Zaanstreek en Waterland met 11,4 ontstane vacatures per 1000 banen, juist omdat deze ligt aan de noordgrens van Amsterdam. Enerzijds kan dat komen doordat veel werknemers in Amsterdam wonen in deze streek, waardoor er relatief veel banen in Zaanstreek en Waterland worden geteld (locatie van een baan wordt toegekend aan de woonplaats van de werknemer in AZW-statistieken). Anderzijds blijkt ook dat een regionale aanbieder die vooral actief is ten noorden van Amsterdam, niet geheel online wordt teruggevonden.

4.4 Aantal vacatures Zorg en Welzijn naar regio



4.5 Vacaturegraad Zorg en Welzijn naar regio



Referenties

UWV (sep 2018). *Hoe werven werkgevers? Inzichten in de zoektocht van bedrijven naar personeel*. URL: https://www.werk.nl/arbeidsmarktinformatie/images/Hoe_werven_werkgevers.pdf.

Colofon

Uitgever

Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Prepress

Centraal Bureau voor de Statistiek

Ontwerp

Edenspiekermann

Inlichtingen

Tel. 088 570 70 70, fax 070 337 59 94
Via contactformulier: www.cbs.nl/infoservice

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen/Bonaire, 2018.
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.