

> TNO rapport voor  
UWV

# Landelijke doorontwikkeling van de UWV Werkverkenner

*Eindrapportage*

# Landelijke doorontwikkeling van de UWV Werkverkenner

## *Eindrapportage*

Rapport voor: UWV

Datum: 21 februari 2018

Auteurs: Dr. E. Dusseldorp (TNO / Universiteit Leiden)  
Drs. H. Hofstetter (TNO)  
Dr. Corine Sonke (NOA)

Met advies van:  
Prof. dr. ir. H.C.W. de Vet (VuMC)  
Prof. dr. S. van Buuren (TNO/Universiteit van Utrecht)  
Prof. dr. R. W. B. Blonk (TNO)  
Prof. dr. R. van den Bergh (NOA)

Projectnummer: 060.02170  
Rapportnummer: R10279

Contact TNO: E. Dusseldorp  
E-mail: [elise.dusseldorp@fsw.leidenuniv.nl](mailto:elise.dusseldorp@fsw.leidenuniv.nl)

**Gezond Leven**  
Schipholweg 77-89  
2316 ZL LEIDEN  
Postbus 3005  
2301 DA LEIDEN  
[www.tno.nl](http://www.tno.nl)

T 088 866 61 00  
[infodesk@tno.nl](mailto:infodesk@tno.nl)

© 2018 TNO

Alle rechten voorbehouden.

Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook, zonder voorafgaande toestemming van TNO.

Indien dit rapport in opdracht werd uitgebracht, wordt voor de rechten en verplichtingen van opdrachtgever en opdrachtnemer verwezen naar de Algemene Voorwaarden voor opdrachten aan TNO, dan wel de betreffende terzake tussen de partijen gesloten overeenkomst.

Het ter inzage geven van het TNO-rapport aan direct belanghebbenden is toegestaan.

Handelsregisternummer 27376655

## Dankwoord

Aan het wetenschappelijke onderzoek naar de UWV Werkverkenner, dat in dit eindrapport wordt beschreven, hebben verschillende mensen een belangrijke bijdrage geleverd. In dit woord vooraf wil ik hen graag bedanken voor hun betrokkenheid, inzet en actieve bijdragen gedurende dit onderzoek dat liep van maart 2014 tot en met december 2017.

Allereerst wil ik de uitvoerende leden van mijn onderzoeksteam heel hartelijk danken:

- > drs. Hedwig Hofstetter (TNO) voor het aanmaken van de juiste variabelen in alle databestanden, voor haar goede organisatie en enorme inzet en voor de basis die zij heeft gelegd voor alle hoofdstukken van deze eindrapportage;
- > dr. Corine Sonke (NOA) voor haar bijdrage aan het gebruikersonderzoek, hoofdstukken 1 en 7 van deze eindrapportage en voor haar geduld en ondersteuning;
- > stagiaire Yi He (van de masteropleiding Statistical Science in the Life and Behavioral Sciences, Leiden University) voor haar bijdrage aan hoofdstukken 4 en 8 van deze eindrapportage en tabellen in de bijlagen;
- > dr. Kim Kranenborg (TNO) voor haar bijdrage aan het gebruikersonderzoek;
- > stagiaire John Zavrakidis (van de masteropleiding Statistical Science in the Life and Behavioral Sciences, Leiden University) voor zijn bijdrage aan de survival analyses (in de 2<sup>e</sup> tussenrapportage);
- > Helga Cebol voor de mooie opmaak van deze eindrapportage en voor haar geduld.

Ook de adviseurs binnen het onderzoeksteam wil ik hier graag bedanken voor hun goede aanvullingen en kritische blik:

- > prof. dr. ir. Riekje de Vet (VuMC);
- > prof. dr. Roland Blonk (TNO);
- > dr. Remko van den Berg (NOA);
- > prof. dr. Stef van Buuren (TNO/Universiteit van Utrecht).

Graag wil ik ook dr. Marta Fiocco bedanken voor de begeleiding van de stagiaires en drs. Sanne Willems voor haar tekstaanvullingen op deze eindrapportage en beiden voor de interessante discussies over survival modellen en logistische regressiemodellen.

Dit onderzoek werd op de voet gevolgd door de leden van de wetenschappelijke begeleidingscommissie, die in vijf bijeenkomsten de wetenschappelijke kwaliteit van het onderzoek gewaarborgd hebben. De discussies tijdens deze bijeenkomsten gaven mede richting aan de vervolgstappen in het onderzoek. Veel dank gaat uit naar de constructieve en gedegen adviezen die zij gaven. De commissie bestond uit:

- > prof. dr. Pierre Koning: bijzonder hoogleraar Arbeidsmarkt en Sociale Zekerheid Vrije Universiteit Amsterdam en universitair hoofddocent Economie Juridische faculteit Universiteit Leiden;
- > dr. Edwin van Hooft: universitair hoofddocent Arbeid en Organisationspsychologie Universiteit van Amsterdam;
- > dr. Wouter van Gils: customer intelligence analist bij UWV divisie Klant & Service (lid t/m januari 2017, wegens baan buiten UWV lidmaatschap overgedragen aan drs. Ewoud van Vegt);
- > drs. Ewoud Vegt: manager Risk & Intelligence bij UWV directie Handhaving (lid vanaf februari 2017).

Het onderzoeksproject is vanuit de directie Strategie, Beleid en Kenniscentrum (SBK) van UWV begeleid door drs. Harriët Havinga, drs. Martijn A. Wijnhoven, drs. Han van der Heul en dr. Maurice Guiaux. Ik dank hen voor de gedegen wijze waarop zij hebben bijgedragen aan de opzet, uitvoering en rapportage van het onderzoek. De vraagstellingen zijn door hen opgesteld in een startnotitie die als leidraad diende voor dit onderzoek. De dataverzameling is door UWV uitgevoerd en met name het toevoegen van de uitkomstmaat (status Werkhervatting op 6, 9 en 12 maanden) was een arbeidsintensief karwei (uitgevoerd door Harriët). Ook zorgde het viertal ervoor om tijdens het onderzoek de focus te houden op de praktische benutting van het wetenschappelijk onderbouwde voorspelmodel van de Werkverkenner in de dienstverlening van UWV. Ik wil specifiek Harriët Havinga bedanken voor de telefoongesprekken (tweewekelijks) waardoor we de koers van het onderzoek konden behouden. Door deze gesprekken werd ik me er van bewust dat regelmatige interactie met de opdrachtgever onmisbaar is voor het slagen van een project. Ook wil ik het hoofd van het kenniscentrum bedanken, dr. Gijsbert van Lomwel, voor zijn fijne, ondersteunende rol op de achtergrond.

Graag bedank ik drs. Brechtje Molenaar, dr. Tineke de Jonge, drs. Nicole Kievits, drs. Paul van Gent en drs. Nienke van der Schaaf van het Werkbedrijf van UWV voor hun bijdragen. Zij toetsten gedurende het onderzoeksproces of de Werkverkenner 'in ontwikkeling' bleef voldoen aan de wensen van de beleidsuitvoering van UWV om werkzoekenden met een verhoogd risico op langdurige werkloosheid in een vroeg stadium extra dienstverlening aan te bieden.

Niet in de laatste plaats een woord van dank aan de UWV-adviseurs en de WW-klienten die in aparte sessies hun aanbevelingen hebben gegeven over de gebruiksvriendelijkheid van de Werkverkenner.

Dr. Elise Dusseldorp, TNO en Universiteit Leiden

Leiden, 10 december 2017

## Samenvatting en aanbevelingen

Dit eindrapport geeft de wetenschappelijke onderbouwing van Werkverkenner 2.0. Hiertoe heeft Toegepast Natuurwetenschappelijk Onderzoeksinstituut (TNO) samen met het Nederlands psychologisch onderzoeks- en adviesbureau (NOA) en het Vrije Universiteit medisch centrum (VUmc) in opdracht van UWV een onderzoek uitgevoerd voor de “Landelijke doorontwikkeling Werkverkenner en validatie, fase 2”. Dit onderzoek liep van maart 2014 tot en met december 2017.

### *Waarom dit onderzoek?*

Eind 2015 is de eerste versie van de Werkverkenner landelijk in gebruik genomen. Dit meetinstrument brengt, aan de hand van een online vragenlijst en registratiegegevens van UWV, de persoons- en situatiekenmerken van recent ingestroomde werklozen met een WW-uitkering in kaart. Op basis hiervan kan UWV de kans op werkherhervatting van individuele werkzoekenden bepalen en dienstverlening op maat aanbieden door te bepalen welke mogelijkheden en belemmeringen bij een werkzoekende beïnvloed kunnen worden om de kans op werk te vergroten. Het onderzoek achter Werkverkenner 1.0 is uitgevoerd onder een WW-instroompopulatie uit 2008-2009 in één regio in Nederland. Nieuw onderzoek was nodig om a) voor zover mogelijk de voorspellende eigenschappen van de Werkverkenner te verbeteren; b) een representatief, landelijk onderzoek uit te voeren op een actueel WW-instroombestand; c) inzicht te verschaffen in de psychometrische eigenschappen van het instrument.

De belangrijkste onderzoeksvragen zijn:

- 1) Wat zijn voorspellende persoons- en situatiekenmerken voor werkherhervatting en duur van de werkloosheid van de WW-instroom?
- 2) Hoe kunnen de onderzoeksresultaten bij vraag 1 vertaald worden naar een objectief scoringssysteem (specificaties) voor een instrument dat direct toepasbaar is in de dienstverlening van UWV?
- 3) Wat zijn de psychometrische eigenschappen van het instrument?

De resultaten worden in deze samenvatting in een andere volgorde besproken dan in het rapport. Hier geven we eerst antwoord op de drie hoofdvragen, daarna bespreken we de vraag of een aparte Werkverkenner nodig is voor specifieke klantgroepen en lichten we de onderzoeksopzet verder toe. Tussen haakjes geven we aan waar in het rapport de resultaten beschreven worden. Aan het eind geven we aanbevelingen voor toekomstig onderzoek.

## Resultaten

### *Voorspellende kenmerken voor werkherhervatting, vraag 1*

Het meest spaarzame, interpreteerbare model om werkherhervatting binnen een jaar te voorspellen bestaat uit 18 factoren (tabel A; zie verder hoofdstuk 3). Dit logistisch regressiemodel voorspelt 70,3% van de werkherhervatting correct (model D in tabel B). In het model komen negen van de 11 voorspellende factoren uit Werkverkenner 1.0 voor. Van de 18 factoren zijn zes harde factoren afkomstig uit de administratie van UWV, worden drie harde en acht zachte factoren uitgevraagd met de online vragenlijst en is één factor (Samenstelling huishouden) een combinatie van vragen uit de online vragenlijst en een administratief gegeven. In tabel A zijn de 18 factoren gepresenteerd in volgorde van hun

unieke bijdrage aan de voorspelling; de factor met de hoogste waarde voor  $\Delta R^2$  (i.e. een maat voor de relatieve, unieke bijdrage aan de voorspelling) staat bovenaan.

#### Voorspelling van werkloosheidsduur

In het onderzoek is ook apart gekeken naar een voorspelmodel voor werkloosheidsduur (oftewel, de tijd tot werkhervatting). Dit is gedaan met het Cox regressiemodel (hoofdstuk 4). De factoren in dit model kwamen grotendeels overeen met de factoren in het meest spaarzame, interpreteerbare logistisch regressie-model (17 van de 20 waren gelijk). De voorspellende eigenschappen van het Cox regressiemodel waren vergelijkbaar met het logistische regressiemodel (zie Model E van tabel B).

UWV is tevens geïnteresseerd in een schatting van de werkloosheidsduur bij aanvang van de WW (per werkzoekende). Aan de hand van de oplossing van het Cox regressiemodel (i.e. de geschatte "survival curve" per individu), hebben we de gemiddelde werkloosheidsduur per individu geschat (i.e. de "median survival time"). Deze voorspelling blijkt zeer onnauwkeurig (de correlatie van de voorspelde duur met de geobserveerde werkloosheidsduur = 0,18). Deze bevinding bevestigt de resultaten van een eerdere studie (Henderson & Keiding, 2005).

Tabel A: Voorspellers van werkhervatting in Werkverkenner 2.0 en Werkverkenner 1.0

	Factoren Werkverkenner 2.0 (in volgorde van bijdrage aan de voorspelling)	Werkverkenner 1.0 (nummer van de factor geeft de volgorde van de bijdrage van de factor aan)	theoretisch model	harde of zachte factor
1.	Leeftijd	1	Wanberg	hard, UWV <sup>1</sup>
2.	Gewerkte jaren in laatste functie	10	Wanberg	hard <sup>2</sup>
3.	Visie op terugkeer naar werk	4	Wanberg	zacht <sup>2</sup>
4.	Soort bemiddelingsberoep	-	Wanberg	hard, UWV <sup>1</sup>
5.	Samenstelling huishouden	-	Wanberg	hard, UWV <sup>3</sup>
6.	Maximale duur WW recht	-	Wanberg	hard, UWV <sup>1</sup>
7.	Bedrijfstak voor WW	-	Wanberg	hard, UWV <sup>1</sup>
8.	Werkzoekgedrag m.b.t. direct contact werkgevers	2	TPB	zacht <sup>2</sup>
9.	Gezondheidsbeleving (zich te ziek voelen; lichamenlijk- & psychisch werkvermogen; ervaren gezondheid)	7. Zich te ziek voelen; 9. Psychisch werkvermogen; 11. Lichamenlijk werkvermogen	Wanberg	zacht <sup>2</sup>
10.	Balans voor- en nadelen niet-werken	-	Motivatie	zacht <sup>2</sup>
11.	WIA 35-min-indicatie	-	Wanberg	hard, UWV <sup>1</sup>
12.	Problemen Nederlandse taal	3	Wanberg	hard <sup>2</sup>
13.	Algemeen werkvermogen	5	Wanberg	zacht <sup>2</sup>
14.	Acceptatiebereidheid voltijd baan	-	Wanberg	zacht <sup>2</sup>
15.	Uren per week in staat te werken	-	Wanberg	zacht <sup>2</sup>
16.	Opleidingsniveau	-	Wanberg	hard, UWV <sup>1</sup>
17.	Acceptatiebereidheid m.b.t. afwijkende tijden	-	Wanberg	zacht <sup>2</sup>
18.	Inkomsten uit werk naast WW	-	Wanberg	hard <sup>2</sup>
	-	6. Werkzoekintentie	TPB	
	-	8. Externe variabele attributie	Motivatie	

<sup>1</sup>Uit de administratie van UWV; <sup>2</sup>Uit de online vragenlijst; <sup>3</sup>combinatie uit de vragenlijst en uit de administratie van UWV;

### Voorspellende eigenschappen verschillende modellen

Om een goed antwoord te geven op de eerste onderzoeksvraag hebben we verschillende voorspelmodellen vergeleken (hoofdstukken 3 en 4). In tabel B zijn de statistische maten van 4 logistisch regressiemodellen (A t/m D) en een Cox regressiemodel (E) weergegeven.

De logistisch regressiemodellen voorspellen werkhervatting binnen 12 maanden, het Cox regressiemodel de werkloosheidsduur. Het meest spaarzame, interpreteerbare logistisch regressiemodel (Model D) blijkt in alle maten beter dan de andere modellen m.u.v. het percentage correct voorspeld. Dat is 0,1%-punt hoger met het Cox regressiemodel. Het meest spaarzame, interpreteerbare logistisch regressiemodel voorspelt goed (70,3% op de test gegevens), past goed ( $R^2 > 0,30$ ), heeft een redelijk onderscheidend vermogen ( $0,70 < AUC < 0,80$ ) en een acceptabele voorspelfout (Brierscore  $< 0,20$ ).

### Beste kandidaat voor Werkverkenner 2.0

Wij raden aan om het meest spaarzame, interpreteerbare logistische regressiemodel te implementeren als Werkverkenner 2.0. Dit logistische regressiemodel voorspelt even goed als het Cox regressiemodel en heeft de volgende praktische voordelen: 1) UWV kan het model gemakkelijk omzetten in een voorspelregel die UWV kan implementeren in zijn systemen; 2) het type voorspelregel is vergelijkbaar met de voorspelregel die UWV gebruikt in Werkverkenner 1.0; en 3) de invloed van de factoren (uitgedrukt als odds ratio) op de kans op werkhervatting is iets makkelijker te interpreteren dan de invloed van de factoren in een Cox regressiemodel (uitgedrukt als hazard ratio).

Tabel B: Overzicht van de statistische maten voor de geschatte modellen voor de voorspelling van werkhervatting / werkloosheidsduur binnen 12 maanden

model	aantal factoren	ontwikkelset		testset	
		Nagelkerke $R^2$ / Cox & Snell $R^2$	AUC index / C-index	Brierscore	percentage correct voorspeld
Model A: Logistisch regressiemodel met alleen intercept (= naïeve methode)	0	-	-	-	51,8
Model B: Logistisch regressiemodel met alleen de factor Leeftijd	1	0,157	0,700	0,220	64,1
Model C: Logistische regressiemodel met factoren van Werkverkenner 1.0	11	0,258	0,751	0,201	68,0
Model D: Meest spaarzame, interpreteerbare logistische regressiemodel	18	0,313	0,778	0,191	70,3
Model E: Meest spaarzame, interpreteerbare Cox regressiemodel	20	0,273	0,704 <sup>b</sup>	0,192 <sup>a</sup>	70,4 <sup>a</sup>

AUC = Area-Under-the-Curve; C-index = Concordance statistic; <sup>a</sup> Bepaald met het Cox regressiemodel voor  $t = 365$ . <sup>b</sup> Deze index is bepaald over een langere periode (t/m april 2016)

### Dienstverlening op maat op basis van Werkverkenner 2.0 in de praktijk, vraag 2

Op basis van de voorspelde kans op werkhervatting binnen 12 maanden besluit UWV om extra dienstverlening in te zetten. Werkzoekenden komen bijvoorbeeld in aanmerking voor een extra gesprek met een adviseur werk wanneer hun voorspelde kans op werkhervatting binnen een jaar lager of gelijk is aan het afkappunt van 0,50. Het afkappunt voor Werkverkenner 2.0 zal door UWV met zorg gekozen worden in overleg met alle betrokken partijen. We raden daarbij aan om niet alleen te kijken naar het hoogste percentage correct voorspeld, maar ook naar het percentage werkzoekenden dat onterecht de extra dienst-

verlening zou worden onthouden bij een gekozen afkappunt (i.e. het percentage fout positieven; zie tabellen 3.5c en 3.5d in hoofdstuk 3).<sup>1</sup>

#### *Diagnose op basis van Werkverkenner 2.0*

We stellen voor om de uitslag per factor van een individuele werkzoekende op de Werkverkenner te presenteren aan adviseurs in vijf kleuren, van rood naar donkergroen, (hoofdstuk 6). De keuze voor vijf kleuren kwam naar voren uit een focusgroep die we hebben gehouden met adviseurs werk. De kleurenschakering geeft per factor weer in hoeverre de score van een werkzoekende belemmerend of bevorderend is voor het hervatten van werk binnen 12 maanden. Om hiervoor een objectief scoresysteem te maken, gebruiken we voor numerieke factoren de groep werkhervatters binnen 12 maanden als normgroep en bepalen we de vijf kleuren op basis van het gemiddelde en de standaarddeviatie in deze groep (d.m.v. afkapwaarden op de *T*-schaal). Voor categorische factoren werken we direct met het criterium status werkhervatting en bepalen we de indeling in kleuren op basis van de odds ratio's in het logistisch regressiemodel met 18 factoren (ontwikkeld in hoofdstuk 3). Op basis van de uiteindelijke visualisatie van de scores van een werkzoekende op de 18 factoren kunnen de adviseurs werk snel een diagnose maken. Het is verder (theoretisch) mogelijk om de toename in kans op werkhervatting te laten zien als de score één of meerdere vakken (i.e. kleuren) opschuift naar rechts (bevorderend). We raden aan om de voorgestelde visualisatie van de scores op de 18 factoren voor de adviseur werk in een pilotonderzoek uit te proberen op bruikbaarheid in de praktijk.

#### *Psychometrische eigenschappen vragenlijst, vraag 3*

De factoren in de Werkverkenner zijn voldoende betrouwbaar voor het doel waarvoor hij ingezet gaat worden (hoofdstuk 7). Bij het bepalen van de psychometrische eigenschappen hebben we gekeken naar betrouwbaarheidsmaten gebaseerd op inter-item relaties (Cronbach's alpha, Guttman's lambda) van de zachte factoren die gemeten zijn door meerdere items in de uitgebreide vragenlijst van Werkverkenner die gebruikt is in dit onderzoek (1.0+). Voor diagnose en begeleiding is de minimaal vereiste betrouwbaarheid 0,70; alle factoren die in Werkverkenner 2.0 voorkomen voldoen aan deze eis (zowel in de totale groep, als in de groepen werkhervatters en niet-werkhervatters). Ook de test-hertest correlaties van de meeste factoren zijn ruim voldoende ( $r \geq 0,70$ ).

De constructvaliditeit van de Werkverkenner is goed. Dit blijkt uit de confirmatieve factoranalyse (CFA): de veronderstelde factorstructuur past goed bij de geobserveerde data. Voor alle items werden factorladingen van voldoende grootte gevonden. Daarnaast is deze factorstructuur ook houdbaar voor verschillende groepen (e.g. de normgroep en de klantgroepen). Ook het CFA model waarbij we de factorladingen en intercepten vastzetten in de verschillende groepen laat een goede fit zien. Dit betekent dat er geen sprake is van item bias.

In het soortgenotenonderzoek van de Werkverkenner vinden we voor enkele schalen (i.e. factoren die door meerdere items worden gemeten) samenhang met inhoudelijk corresponderende schalen uit de Arbeidsmarktmotivatie en Werkzoekgedrag Vragenlijst (AWV); met name tussen de factor Visie op terugkeer naar werk, de belangrijkste zachte factor in Werkverkenner 2.0, en de factor Inschatting kans op werk ( $r = 0,69$ ). Voor de

---

<sup>1</sup> Hoe zorgvuldig het afkappunt ook gekozen wordt, bij elk afkappunt zijn er werkzoekenden die niet (meteen) in aanmerking komen voor de extra dienstverlening van UWV. UWV heeft hiervoor een vangnet ingericht. UWV nodigt deze werkzoekenden uit voor extra dienstverlening als blijkt dat ze na een aantal maanden nog een WW-uitkering hebben.



meeste schalen vinden we echter een minder hoge samenhang dan we van tevoren hadden verwacht. De AWW sluit niet op alle gebieden aan bij de Werkverkenner. Daarnaast zijn de Werkverkenner en de AWW niet op dezelfde dag ingevuld, wat een invloed kan hebben gehad op de samenhang tussen de schalen. Op basis van het soortgenoten onderzoek is er geen reden om schalen van de Werkverkenner niet mee te nemen in het voorspelmodel voor werkhervatting.

Om de mate van sociaal wenselijkheid te onderzoeken hebben we, gegeven de onderzoeksopzet van de huidige studie, ons beperkt tot het inspecteren van de verdelingen van de antwoorden op de items. Deze geven geen indicatie voor een hoge mate van sociaalwenselijke beantwoording. Een geringe mate van sociaalwenselijke beantwoording doet geen afbreuk aan de kwaliteit van de voorspelregel voor werkhervatting. De voorspelkracht van de regel zal niet afnemen, doordat het aannemelijk is dat de mate van sociaalwenselijkheid in de antwoorden van de huidige steekproef niet zal verschillen van de mate van sociaalwenselijkheid van toekomstige werkzoekenden.

#### *Aparte Werkverkenner voor specifieke klantgroepen niet nodig*

In het onderzoek hebben we voorspelmodellen ontwikkeld voor drie door UWV aangegeven specifieke WW-klantgroepen met een grote afstand tot de arbeidsmarkt (hoofdstuk 8). De drie klantgroepen zijn: mensen van 50 tot en met 54 jaar, 55-plussers en mensen met een WIA 35-min-indicatie. Per klantgroep hebben we de voorspellende eigenschappen van het klant-specifieke model vergeleken met de voorspellende eigenschappen van het model voor de totale onderzoeksgroep. De voorspellende eigenschappen van klant-specifieke modellen bleken vrijwel gelijk aan die van Werkverkenner 2.0. Er is dus geen betere predictie accuraatheid verkregen door een voorspelmodel te gebruiken dat specifiek is afgestemd op de klantgroep. Met andere woorden, voor deze klantgroepen kunnen we volstaan met de voorspelling door Werkverkenner 2.0.

#### *Werkzoekenden tevreden over gebruiksvriendelijkheid Werkverkenner*

De acht werkzoekenden die deelnamen aan een kwalitatief gebruikersonderzoek naar de Werkverkenner waren over het algemeen positief (hoofdstuk 6). Het starten en invullen van Werkverkenner 1.0+ gaat vlot en duurt ongeveer 7-15 minuten. Dit wordt als kort ervaren door de deelnemers. Hierbij dient te worden opgemerkt dat Werkverkenner 1.0+ de meest uitgebreide versie van de Werkverkenner is met 50 vragen aan de klant<sup>2</sup>. Het is dus positief dat deze lange versie als kort wordt ervaren. Ook de taken, zoals het bedienen van en het navigeren in de vragenlijst en het zoeken naar hulp, verlopen over het algemeen zonder problemen en worden als prima, simpel, of helder benoemd. Gemiddeld beoordeelden de deelnemers de Werkverkenner met het rapportcijfer 7,6. De deelnemers vinden de vragen niet kwetsend. Ze vinden de instructie, de vragen en de antwoordcategorieën duidelijk. Het taalgebruik vindt men eenvoudig. Verder worden de leesbaarheid en de lettergrootte van de teksten als goed ervaren. De vormgeving van de UWV website wordt als goed tot voldoende beoordeeld. De UWV huisstijl ziet er volgens de deelnemers betrouwbaar en professioneel uit. De Werkverkenner is onderdeel van de werkmap en is beveiligd via DigiD. Dit wordt als veilig ervaren en men heeft hier voldoende vertrouwen in. UWV wordt ook gezien als een organisatie die te vertrouwen is. Op basis van deze resultaten kunnen we concluderen dat de Werkverkenner door de deelnemers als gebruiksvriendelijk wordt ervaren. Wel kwam voren dat het voor deelnemers soms niet duidelijk is waarom zij de Werkverkenner moeten

---

<sup>2</sup> Deze uitgebreide versie is aan de hand van de resultaten van het huidige onderzoek ingekort tot 2.0. Het voorspelmodel voor Werkverkenner 2.0 bestaat uit 18 voorspellende factoren voor werkhervatting, waarvan 12 factoren worden vastgesteld met 23 vragen aan de klant (zie ook tabel A).

invullen, wat het nut voor henzelf is en wat ze kunnen verwachten van UWV. Ook kwam naar voren dat de deelnemers behoefte hebben aan tips die beter aansluiten bij hun specifieke situatie (meer maatwerk).

#### *Adviseurs werk hebben behoefte aan meer informatie bij gebruik diagnose*

Eerder gaven we al aan dat de adviseurs van UWV die deelnamen aan de focusgroep meer behoefte hebben aan nuancering in kleuren dan de huidige rood en groen in de diagnose van de Werkverkenner. In hetzelfde kwalitatieve onderzoek gaven ze aan dat ze mogelijkheden willen hebben om extra resultaten en informatie te kunnen bekijken van de klant (bijvoorbeeld door te klikken op een i-button). Ook is behoefte aan handvaten voor acties naar aanleiding van de resultaten (bijv. beeldvragen of verwijzing naar e-producten).

#### *Schatting werkhervatting per regionale koplopervestiging hetzelfde*

Een belangrijk resultaat van het onderzoek is dat regionale koplopervestiging geen voorspellende factor is in Werkverkenner 2.0 (hoofdstuk 3). Tijdens het ontwikkelen van de voorspelmodellen hebben we de factor Regio koplopervestiging in de analyses meegenomen. Deze factor viel echter steeds uit het model tijdens de analyse-stappen waarin we het meest spaarzame model selecteerden (m.u.v. het model specifiek voor de klantgroep 55-plussers). Het wegvallen van deze factor betekent dat de regionale koplopervestiging geen unieke bijdrage levert aan de voorspelling van werkhervatting (bovenop wat al verklaard wordt door de andere factoren in Werkverkenner 2.0). Het is dus geen belangrijke voorspeller gebleken.

Het is aannemelijk dat de factoren Bemiddelingsberoep soort en Bedrijfstak voor WW die wel in Werkverkenner 2.0 zitten de regionale verschillen tussen de koplopervestigingen in kans op werkhervatting genoeg verklaren. We kunnen hiermee echter niet beweren dat de relaties tussen de factoren van Werkverkenner 2.0 en werkhervatting in iedere regio gelijk zijn (i.e. interactie-effecten tussen Regio koplopervestiging en factoren in WW 2.0 zijn niet onderzocht). Ook kunnen we de resultaten niet zonder meer generaliseren naar een regio in zijn geheel, aangezien in het huidige onderzoek per regio slechts één UWV-vestiging (i.e. de koplopervestiging) heeft meegedaan. We hebben gezien dat de koplopervestigingen landelijk representatief zijn, maar per regio is de representativiteit niet onderzocht.

## Onderzoeksopzet

### *Methode*

De opzet bestond uit een prospectief cohort onderzoek. Hiervoor hebben wij aan werkzoekenden een uitgebreide onderzoeksversie van de Werkverkenner (1.0+) aangeboden aan het begin van hun werkloosheid. De Werkverkenner werd ten tijde van het onderzoek tussen de zesde en tiende week na toekenning van de WW-uitkering aangeboden. De werkzoekenden hadden vervolgens twee weken de tijd om de vragenlijst in te vullen. Van de werkzoekenden is vervolgens de status werkhervatting binnen 6, 9 en 12 maanden na aanvang van het WW-recht bepaald op basis van gegevens uit het WW-systeem en het systeem Polis. De werkloosheidsduur is vastgesteld tot en met het einde van de dataverzameling op 30 april 2016.

### *Steekproef, respons en representativiteit*

De onderzoekspopulatie betreft werkzoekenden binnen de WW-instream die gebruik maken van de digitale werkmap van UWV en langer dan drie maanden recht hebben op een WW-uitkering. De steekproef bestaat uit werkzoekenden die stonden ingeschreven op één van de 11 koplopervestigingen van UWV (verspreid over het hele land) en de WW-uitkering

toegekend kregen tussen 1 maart 2014 en 28 februari 2015 (dus een cohort van een jaar). Van de 76.817 werkzoekenden uit de steekproef, hebben 53.238 werkzoekenden de vragenlijst van Werkverkenner 1.0+ ingevuld. De respons is 69,3%. Deze 53.238 werkzoekenden verschillen nauwelijks van de non-respons groep op achtergrondkenmerken, zoals leeftijd, geslacht, opleiding, nationaliteit, bedrijfstak voor WW. De gevonden effectgroottes zijn zeer klein tot klein. We concluderen dat de respons groep geen selecte groep is.

Slechts 0,3% van de werkzoekenden heeft missende waarden op één of meerdere variabelen. Het gaat hier om variabelen waarvan de waarden (nog) niet in de administratie van UWV geregistreerd waren (zoals opleiding en beroep). In totaal zijn 53.079 werkzoekenden beschikbaar voor de analyses (i.e. netto steekproef). De netto steekproef verschilt niet qua achtergrondkenmerken van werkzoekenden van niet-koplopervestigingen en de non-responsgroep. We concluderen dat de geanalyseerde steekproef representatief is voor de landelijke WW-instroompopulatie die voldoet aan de indicatiecriteria voor de Werkverkenner.

#### *Statistische analyse*

Voor het ontwikkelen van de voorspelmodellen hebben we de totale steekproef werkzoekenden ( $n = 53.079$ ) aselekt verdeeld in een ontwikkelset ( $n = 26.541$ ; 50%), validatieset ( $n = 13.269$ ; 25%) en testset ( $n = 13.269$ ; 25%). De ontwikkelset gebruiken we om het voorspelmodel te schatten, de validatieset gebruiken we voor modelselectie (i.e. het reduceren van aantal factoren in het model) en de testset gebruiken we voor het bepalen van de voorspellende eigenschappen van het model. Het voordeel van deze werkwijze is dat we op een zo goed mogelijke manier de voorspelkracht van het model kunnen bepalen voor toekomstig werkzoekenden. In de testset zitten namelijk de werkzoekenden die niet gebruikt zijn voor het ontwikkelen van het voorspelmodel. Op deze manier bootsen we “toekomstige werkzoekenden” na en kunnen we de verwachte (gegeneraliseerde) voorspelfout bepalen en het verwachte onderscheidend vermogen.

Voor het voorspellen van werkhervatting hebben we logistische regressieanalyses uitgevoerd. We hebben ook alternatieve methoden onderzocht (o.a. random forest, elastic net, support vector machine). Van deze alternatieve methoden bleek alleen random forest iets beter dan logistische regressie (1%-punt hoger in percentage correct voorspeld). Het random forest model leverde echter een minder goed interpreteerbaar model op dan het logistische regressiemodel. Voor het voorspellen van werkloosheidsduur hebben we Cox regressieanalyse toegepast, hiermee kunnen we ook de kans op werkhervatting bepalen op een bepaald tijdstip ( $t$ ). De kwaliteit van de verschillende voorspelmodellen in tabel B hebben we vergeleken met behulp van de statistische maten in tabel C.

Voor het beantwoorden van de derde onderzoeksvraag hebben we de betrouwbaarheid bepaald met het volledige onderzoekbestand en met een test-hertest-bestand ( $n = 1620$ ). De werkzoekenden in het test-hertest-bestand hebben de vragenlijst van de Werkverkenner twee keer ingevuld met een tussenperiode van minimaal twee tot maximaal vier weken. De constructvaliditeit hebben we bepaald met het volledige onderzoekbestand en met een soortgenotenbestand ( $n = 619$ ). De werkzoekenden in het soortgenotenbestand hebben vlak na elkaar de vragenlijst van de Werkverkenner ingevuld en de Arbeidsmarktmotivatie en Werkzoekgedrag Vragenlijst (AWV).

Om de gebruiksvriendelijkheid van de Werkverkenner te meten is kwalitatief onderzoek verricht. Met een aantal werkzoekenden met een WW-uitkering is een user-walkthrough gehouden en met adviseurs werk van UWV is een focusgroep sessie gehouden.

Tabel C: Overzicht van de gebruikte statistische maten in dit onderzoek

maat	betekenis	richtlijn	op welke dataset
Nagelkerke $R^2$ / Cox & Snell $R^2$	Geeft aan hoe goed het logistisch regressiemodel (Nagelkerke $R^2$ ) of Cox regressiemodel (Cox & Snell $R^2$ ) past bij de data	$R^2 \geq 0,30$ is goed passend	ontwikkelset
$\Delta R^2$	Relatieve toename in Nagelkerke $R^2$ (of Cox & Snell $R^2$ voor Cox regressie); een maat voor de unieke bijdrage van een factor aan de voorspelling (in vergelijking met de andere factoren)	Een bijdrage van $\Delta R^2 > 0,001$ is aan te raden; een lagere waarde dan 0,001 betekent dat een factor een erg kleine unieke bijdrage levert aan de voorspelling	ontwikkelset
Area-under-the-Curve (AUC) of C-index	Het onderscheidend vermogen van het model	0,70 < AUC < 0,80 is redelijk; AUC $\geq 0,80$ is goed	validatieset en testset
Brierscore	Gemiddelde voorspelfout van het model	Brierscore < 0,20 is goed	validatieset en testset
Sensitiviteit	Het percentage correct voorspeld binnen de groep werkhervatters oftewel het percentage terecht positieven	Afhankelijk van de context	testset
Specificiteit	Het percentage correct voorspeld binnen de groep niet-werkhervatters oftewel het percentage terecht negatieven	Afhankelijk van de context	testset
Percentage correct voorspeld	Het percentage correct voorspeld binnen de totale groep	Voor vergelijking van de modellen wordt het hoogste percentage gebruikt; in de praktijk is de keuze van het percentage afhankelijk van de context (beleid, budget en arbeidsmarktontwikkelingen)	testset

### Aanbevelingen voor toekomstig onderzoek

Om de voorspelkracht van de Werkverkenner in de toekomst te blijven toetsen en verbeteren bevelen we externe validatie aan. Met behulp van een nieuwe actuele dataset wordt het ontwikkelde model dat ten grondslag ligt aan Werkverkenner 2 opnieuw getoetst en zo nodig geactualiseerd (bijv. door de gewichten aan te passen). UWV is voornemens om externe validatie onderdeel te laten uitmaken van het op te zetten kwaliteitsprogramma periodiek onderhoud en doorontwikkeling Werkverkenner.

De theoretische modellen waarop de Werkverkenner is gebaseerd, veronderstellen een onderlinge relatie tussen verschillende factoren uit de Werkverkenner. Volgens de Theory of Planned Behavior (TPB; Ajzen, 1985), bijvoorbeeld, wordt werkzoekgedrag voorspeld uit werkzoekintentie en wordt werkzoekintentie op haar beurt voorspeld door factoren zoals attitude, subjectieve norm en self-efficacy (perceived behavioral control). Dit impliceert dat het effect van de factor attitude op werkzoekgedrag gemedieerd wordt door de factor werkzoekintentie (zie bijv. Van Hoof, Born, Taris, Van der Flier, & Blonk, 2004). In het uiteindelijke voorspelmodel van Werkverkenner 2.0 is alleen de factor werkzoekgedrag van de TPB opgenomen, de rest van de factoren van de TPB zijn weggefallen (zie tabel A). Dit resultaat is theoretisch te verklaren volgens de TPB omdat werkzoekgedrag de meest proximale voorspeller is van werkhervatting; de andere factoren van de TPB beïnvloeden weer de factor werkzoekgedrag (en worden ook wel distale factoren genoemd). In het huidige onderzoek hebben we geen rekening gehouden met deze onderliggende structuur

van de factoren van de TPB. We hebben er verder voor gekozen om factoren die verbanden lieten zien die tegenstrijdig waren met de achterliggende theorie te verwijderen uit het model. Echter, mogelijk zijn onderliggende mediatie-effecten een oorzaak van deze tegenstrijdige verbanden. We raden daarom aan om deze relaties verder te onderzoeken met zogenaamde padmodellen. Belangrijke onderzoeksvragen die hierbij een rol spelen zijn: levert een padmodel een betere voorspelling op en/of een betere indicatie voor dienstverlening?

Naast mediatie-effecten spelen mogelijk ook moderator-effecten (i.e. interacties) een rol in de voorspelling van werkhervatting. Bijvoorbeeld, oudere werkzoekenden maken doorgaans meer realistische inschattingen (bijv. of ze snel het werk zullen hervatten) dan jongere werkzoekenden. Dit impliceert een interactie-effect tussen Leeftijd en Visie op terugkeer naar werk. Het totaal aantal factoren in de Werkverkenner was in dit onderzoek te groot om exploratieve analyses uit te voeren naar interactie-effecten (i.e. alle mogelijke interactie-effecten op te nemen in het model). We bevelen voor toekomstig onderzoek aan om theoretisch onderbouwde interactie-effecten mee te nemen in het voorspelmodel om op deze wijze mogelijk de voorspelling te verbeteren.

Sommige factoren uit de Werkverkenner hebben een dynamisch karakter (bijvoorbeeld Werkzoekintentie en Self-efficacy). Dat wil zeggen dat de scores op deze factoren fluctueren over de tijd. In een recente studie is aangetoond dat met name de verandering over de tijd in self-efficacy voorspellend is voor werkhervatting (Lagerveld, Brenninkmeijer, Blonk, Twisk & Schaufeli, 2016). Interessant voor toekomstig onderzoek is om deze veranderingen over tijd mee te nemen in de voorspelling van werkhervatting. Ook is het interessant om te onderzoeken in hoeverre UWV haar dienstverlening beter op maat kan aanbieden als dynamische factoren vaker worden uitgevraagd.

De psychometrische eigenschappen die we hebben aangetoond in het huidige onderzoek zijn gebaseerd op de uitgebreide vragenlijst Werkverkenner 1.0+. We kunnen niet vanzelfsprekend aannemen dat de gevonden betrouwbaarheden en construct validiteit ook gelden voor Werkverkenner 2.0. Er kunnen bijvoorbeeld volgorde effecten van de vragen optreden. We raden daarom aan om de analyses voor de betrouwbaarheden (gebaseerd op inter-item relaties) en voor de constructvaliditeit (confirmatieve factoranalyse) te herhalen voor Werkverkenner 2.0 als deze in gebruik genomen gaat worden.

De user-walkthrough is een vorm van kwalitatief onderzoek waarbij, in dit geval, met een kleine groep WW-klienten uitgebreid is ingegaan op de ervaringen met de Werkverkenner. Overwogen kan worden aanpassingen naar aanleiding van de user-walkthrough in een grotere, representatieve groep respondenten te testen. Dit zou, bijvoorbeeld, gedaan kunnen worden met een feedback vragenlijst na het aanpassen van de introductie of tips op maat. Ook wordt geadviseerd om nog een user-walkthrough te houden met klantgroepen die nu wat onderbelicht zijn gebleven in dit gebruikersonderzoek, bijvoorbeeld lager opgeleiden, mensen met een niet-Nederlandse nationaliteit en minder digivaardigen.

### *Tot slot*

De doelstellingen van het onderzoek naar de Landelijke doorontwikkeling van de Werkverkenner zijn behaald. Ten eerste hebben we de voorspellende eigenschappen van de Werkverkenner kunnen verbeteren, zij het in bescheiden mate: van 69% correct voorspeld door Werkverkenner 1.0 destijds in de steekproef van Noord-Holland (van 2008 - 2009) naar 70% correct voorspeld door Werkverkenner 2.0 in de testset van de huidige, omvangrijkere steekproef (van 2014 - 2015). Ten tweede is de huidige steekproef representatief gebleken voor de landelijke WW-instroom. Ten slotte hebben we inzicht gegeven in de psychometrische eigenschappen van het instrument gebaseerd op de uitgebreide vragenlijst die is afgenomen in dit onderzoek (Werkverkenner 1.0+).

# Inhoudsopgave

Dankwoord .....	i
Samenvatting en aanbevelingen.....	iii
1 Inleiding .....	1
1.1 De ontwikkeling van de Werkverkenner: Beschrijving van het voorafgaande onderzoek.....	1
1.1.1 De Werkverkenner .....	1
1.1.2 Theoretische onderbouwing.....	2
1.1.3 Werkverkenner 1.0 .....	3
1.1.4 Op weg naar Werkverkenner 2.0 .....	6
1.2 Het huidige onderzoek.....	7
2 Beschrijving van de onderzoeksgroep .....	9
2.1 Beschrijving van de onderzoekspopulatie .....	9
2.2 Selectie van werkzoekenden uit steekproef.....	11
2.3 Beschrijvende kenmerken van de steekproef .....	11
3 Resultaten voorspelling werkhervatting .....	19
3.1 Onderzoeksvraag 1 en deelvragen .....	19
3.2 Status werkhervatting .....	19
3.3 Naïeve voorspelmethode.....	20
3.4 Invloed van iedere factor apart op status werkhervatting .....	20
3.4.1 Voorspellende eigenschappen van het model met alleen de factor Leeftijd .....	23
3.5 Voorspelling werkhervatting met meerdere factoren samen .....	25
3.5.1 Resultaten stap 1: hoe voorspellend zijn alle factoren samen .....	26
3.5.2 Resultaten stap 2: Het beter interpreteerbare model .....	29
3.5.3 Resultaten stap 3: Het meest spaarzame, interpreteerbare model .....	29
3.6 Alternatieve methoden.....	38
3.7 Logistische regressie met factoren Werkverkenner 1.0 .....	38
4 Resultaten voorspelling werkloosheidsduur.....	41
4.1 Voorspelling van werkloosheidsduur met alle factoren samen.....	42
4.2 Voorspelling van werkloosheidsduur met alleen interpreteerbare factoren .....	43
4.3 Voorspelling van werkloosheidsduur met meest spaarzame, interpreteerbare model.....	44
4.4 Schatting van gemiddelde werkloosheidsduur per persoon .....	49
4.5 Factoren en voorspelmodel Werkverkenner 2.0 .....	50
5 Respons en representativiteit .....	52
5.1 Respons .....	52
5.1.1 Resultaten responsanalyse.....	53
5.2 Representativiteitsanalyse .....	56
5.2.1 Resultaten representativiteitsanalyse van de koplopervestigingen .....	56
5.2.2 Resultaten representativiteitsanalyse van de geanalyseerde onderzoeksgroep .....	59
6 Diagnose: normering en gebruik.....	62
6.1 Normering.....	63
6.1.1 Normeringsmethode .....	63
6.1.2 Resultaten van de normeringsschalen.....	69
6.1.3 Impact van beïnvloeding van een factor op de kans op werkhervatting .....	75
6.2 Gebruik van de Werkverkenner in de praktijk .....	77



6.2.1	Gebruikersonderzoek onder WW-klienten .....	78
6.2.2	Gebruikersonderzoek onder adviseurs werk.....	84
7	Psychometrische eigenschappen .....	91
7.1	Betrouwbaarheid .....	94
7.1.1	Maten van betrouwbaarheid gebaseerd op inter-item relaties .....	95
7.1.2	Test-hertest betrouwbaarheid .....	100
7.2	Construct validiteit .....	108
7.2.1	Confirmatieve factoranalyse voor de totale groep .....	108
7.2.2	Invariantie factor structuur en mogelijke item bias bij verschillende subgroepen .....	112
7.2.3	Soortgenoten onderzoek .....	114
7.3	Sociaal wenselijke beantwoording .....	127
7.4	Predictieve validiteit .....	128
8	Voorspelmodel per klantgroep .....	130
8.1	Methode per klantgroep .....	132
8.2	Resultaten per klantgroep .....	133
8.2.1	Resultaten voor de klantgroep WW 50 t/m 54 jaar .....	133
8.2.2	Resultaten voor de klantgroep WW 55+ .....	136
8.2.3	Resultaten voor de klantgroep WW 35-min .....	137
9	Referenties .....	139

# 1 Inleiding

Per 1 augustus 2014 is het onderzoek “Landelijke doorontwikkeling Werkverkenner en validatie” van start gegaan. In opdracht van UWV wordt dit onderzoek uitgevoerd door TNO in samenwerking met NOA en het VUmc. Het belangrijkste doel van het onderzoek is de doorontwikkeling van een screenings- en diagnose-instrument voor de WW-klant tot Werkverkenner 2.0. Dit doel zal worden gerealiseerd door:

- 1) de voorspellende eigenschappen van de bestaande Werkverkenner 1.0 te verbeteren;
- 2) het instrument meer geschikt te maken voor landelijk gebruik (door het gebruik van een grote representatieve steekproef) en te actualiseren (door gebruik te maken van een zo recent mogelijke WW-instroompopulatie);
- 3) de psychometrische eigenschappen te bepalen van de Werkverkenner;
- 4) te onderzoeken of de voorspellers voor werkhervatting hetzelfde zijn in verschillende klantgroepen van UWV (bijv. arbeidsbeperkten of 50-plussers).

De Werkverkenner is een meetinstrument dat gebruik maakt van een online vragenlijst die is ontwikkeld door UWV voor mensen die recht hebben op een werkloosheidsuitkering (WW). De werkzoekende krijgt de Werkverkenner aangeboden via UWV en vult deze (vrijwillig) in aan het begin van de werkloosheid. Naast de gegevens uit de online vragenlijst worden aan de Werkverkenner ook gegevens toegevoegd van de klant die UWV al vastgelegd heeft in de registratie. De belangrijkste doelen van de Werkverkenner zijn 1) **Dienstverlening op maat**: Op basis van de resultaten van de Werkverkenner wordt de kans op werkhervatting bepaald. Het gaat hierbij om de kans van het individu (de werkzoekende) op het hervatten van werk. Deze kans wordt gebruikt om de best passende en door UWV betaalbare vorm van dienstverlening te bepalen voor de werkzoekende; 2) **Diagnose**: Op basis van de scores op de factoren van de Werkverkenner wordt een eerste diagnose gemaakt van de mogelijkheden en belemmeringen die een werkzoekende heeft om werk te vinden. De diagnose geeft inzicht in welke factoren beïnvloed kunnen worden om de kans op werk te vergroten. De adviseurs van UWV gebruiken de diagnose als kapstok voor het gesprek met de klant.

Een uitgebreide versie van de Werkverkenner (1.0+, zie bijlage A) is vanaf 1 maart 2014 tot en met 28 februari 2015 op 11 koplopervestigingen van UWV ingevoerd. Deze uitgebreide versie wordt gebruikt in het huidige onderzoek. In de volgende paragrafen zullen we nader ingaan op (de geschiedenis van) de Werkverkenner en het huidige onderzoek.

## 1.1 De ontwikkeling van de Werkverkenner: Beschrijving van het voorafgaande onderzoek

### 1.1.1 De Werkverkenner

De Werkverkenner brengt de persoons- en situatiekenmerken van werklozen in de WW in kaart d.m.v. vragen en stellingen naar harde factoren (bijv. jaren werkervaring) en zachte factoren (zoals gezondheidsbeleving of visie op terugkeer naar werk), die uit onderzoek naar voren zijn gekomen als voorspellend voor het vinden van werk. Werkverkenner 1.0 is het resultaat van drie stadia van wetenschappelijk onderzoek uitgevoerd door de Rijksuniversiteit Groningen (RUG).

Ten eerste is er een literatuuronderzoek uitgevoerd naar voorspellende factoren van werkhervatting. Vervolgens hebben een cross-sectioneel en een longitudinaal onderzoek plaatsgevonden. De resultaten van deze drie onderzoeken hebben geleid tot de 11 factoren die deel uitmaken van Werkverkenner 1.0. We zullen in de volgende twee paragrafen nader ingaan op deze voorafgaande onderzoeken.

### 1.1.2 *Theoretische onderbouwing*

In deze paragraaf zal een samenvatting worden gegeven van het literatuuronderzoek dat de basis heeft gevormd voor de ontwikkeling van de Werkverkenner. Voor een uitgebreidere beschrijving van het literatuuronderzoek wordt verwezen naar Schellekens, Abma, Nijland, Postma, van Bebber, Mulders, Havinga, & van Brakel (2007) en Brouwer, Schellekens, Bakker, Verheij, Steegen, Havinga, & van Brakel (2011).

De literatuurstudie richtte zich op persoons- en situatietekenen van werklozen die mogelijk voorspellend zijn voor de kansen op werkhervatting. Uit een meta-analyse kwam naar voren dat er een verband is tussen persoonlijke kenmerken en werkzoekgedrag en een verband tussen werkzoekgedrag en het vinden van een baan (Kanfer, Wanberg & Kantrowitz, 2001). Voor het doel werkhervatting blijken twee aspecten van belang, namelijk “kunnen” en “willen”. Onder het “kunnen” vallen kennis en vaardigheden, maar ook omgevings- en situatiefactoren die een faciliterende of remmende invloed kunnen hebben (Ford, 1992). Of iemand ook handelt naar zijn “kunnen” wordt bepaald door de perceptie van het eigen kunnen (Baumeister, Campbell, Krueger, & Vohs, 2003). “Willen” wordt hierbij gezien als motivatie voor een bepaald doel en wordt beïnvloed door de aantrekkelijkheid van het doel en de verwachting het doel te kunnen bereiken (Vroom, 1964). Zowel bij “kunnen” als “willen” spelen de perceptie van de eigen mogelijkheden en de mogelijkheden die de situatie biedt een belangrijke rol.

Naar aanleiding van de literatuurstudie door Schellekens et al. (2007) zijn drie theoretische modellen als uitgangspunt genomen voor de ontwikkeling van de Werkverkenner. De modellen bevatten harde en zachte persoons- en situatietekenen, die relevant lijken in relatie tot werkhervatting (Schellekens et al., 2007; Brouwer et al., 2011). Dit zijn het multidisciplinaire model van Wanberg (Wanberg, Song & Hough, 2002), de theorie van gepland gedrag (Theory of Planned Behaviour/TPB; Ajzen, 1991) en het motivatiemodel van Vroom (Vroom, 1964).

Het eerste theoretische model is het multidisciplinaire model van Wanberg (Wanberg et al., 2002). Dit model omvat, behalve psychologische componenten, ook economische en sociologische componenten. Daarnaast zijn ook contextfactoren zoals de arbeidsmarkt onderdeel van dit model. De zeven groepen van variabelen die volgens dit model van invloed zijn op de kansen op een succesvolle werkhervatting van werklozen zijn: arbeidsmarkt, menselijk kapitaal, sociaal kapitaal, obstakels, economische noodzaak, werkzoekgedrag en discriminatie.

Bij de component arbeidsmarkt gaat het om de vraag die er op de arbeidsmarkt is naar werknemers met bepaalde soorten arbeidskwalificaties. Dit wordt geoperationaliseerd door te vragen naar het beeld dat een werkzoekende heeft van zijn kansen op de arbeidsmarkt. Met de component menselijk kapitaal worden de arbeidskwalificaties van een werkzoekende bedoeld. Dit omvat de competenties, de opleiding, persoonlijkheid en andere persoonlijke kenmerken van de werkzoekende (Wanberg et al., 2002). De component sociaal kapitaal heeft betrekking op het netwerk van de werkzoekende dat op verschillende manieren (instrumenteel, informatief en/of emotioneel) ondersteunend kan zijn bij het zoeken naar

werk. Kenmerken en omstandigheden van de werkzoekende en de omgeving die het vinden van werk bemoeilijken, vormen de component obstakels. Gezondheidsproblemen vallen hieronder, maar bijvoorbeeld ook de zorg voor jonge kinderen met name bij vrouwen (Van der Heul, Berendsen, Van de Eijken, & Vlek, 2003; Van der Heul, 2006). Andere voorbeelden van obstakels zijn de beschikbaarheid voor afwijkende werktijden of voldoende werkuren (Wanberg et al., 2002). De component Economische noodzaak heeft betrekking op de invloed van financiële factoren op het werkzoekgedrag en daarmee op de kans op werk. Ook de duur van het recht op een WW-uitkering valt onder deze component. Het belang van de uitkeringsduur op de kans op werkhervatting wordt benadrukt in de economische literatuur (e.g., De Groot & Van der Klaauw, 2014)<sup>3</sup>. Een belangrijk onderscheid bij de component werkzoekgedrag is de intensiteit (frequentie, aantal uren) en de kwaliteit (voorbereiding sollicitatie, cv) van het zoeken naar werk. Daarnaast is ook van belang hoe duidelijk de werkzoekende een beeld heeft van het soort werk dat hij zoekt. De laatste component discriminatie, omvat oneigenlijke selectiecriteria van werkgevers op niet-werkgerelateerde eigenschappen als leeftijd, geslacht en afkomst.

Het tweede theoretische model is de Theory of Planned Behaviour (TPB) (Ajzen, 1991). Dit model beschrijft de factoren die van invloed zijn op het wel of niet vertonen van bepaald gedrag. Omdat uit onderzoek bleek dat werklozen die meer werkzoekgedrag vertonen sneller werk vinden (Kanfer et al., 2001), is de TPB vertaald naar de context van werkzoekgedrag. Volgens het TPB model is (werkzoek)intentie de belangrijkste voorspeller voor (werkzoek)gedrag en wordt werkzoekintentie zelf weer beïnvloed door drie factoren: subjectieve norm, attitude en self-efficacy. Met subjectieve norm wordt (in het algemeen) de sociale druk bedoeld die iemand ervaart om een bepaald gedrag uit te voeren. Positieve druk heeft een positieve invloed op de intentie om het gedrag uit te voeren (en omgekeerd heeft negatieve druk een negatieve invloed). Attitude betreft de positieve of negatieve houding van iemand ten opzichte van bepaald gedrag. Een positieve houding heeft een positieve invloed op de intentie. Self-efficacy, tenslotte, is het vertrouwen dat men in zichzelf heeft dat men succesvol het gedrag kan uitvoeren. Een hogere mate van self-efficacy vergroot de intentie.

Het derde theoretische model, het motivatiemodel (Vroom, 1964), heeft betrekking op het eerder genoemde aspect “willen”. Volgens dit model wordt motivatie bepaald door de aantrekkelijkheid van het doel (werk) en verwachtingen over de haalbaarheid van het doel. Ook de manier waarop men succes en mislukking verklaard (attributie) is van invloed op motivatie (Weiner, 1980; 1986). In het vooronderzoek is daarbij specifiek gekeken naar de houding ten opzichte van werk (werkattitude), de motivatie voor werk en zelfvertrouwen.

Uit de literatuur kwam tenslotte nog het belang van “zachte” competenties (zoals sociale vaardigheden en stressbestendigheid) naar voren. Bij gelijke geschiktheid zijn het deze competenties die bepalen wie het beste past bij een functie (Schellekens et al, 2007; Brouwer et al., 2011).

### 1.1.3 *Werkverkenner 1.0*

Op basis van de literatuurstudie zijn mogelijke voorspellers van werkhervatting geselecteerd op het gebied van persoons- en werkgegevens, factoren van de thuissituatie en arbeidsverleden en zachte factoren gerelateerd aan persoonlijkheid, gedrag en gezondheid.

---

<sup>3</sup> Een interessante bevinding in De Groot & Van der Klaauw (2014) is dat het inkorten van de uitkeringsduur zorgde voor het sneller hervatten van het werk, echter het leidde ook tot een verslechtering van de kwaliteit van het werk (e.g., meer slecht betaalde banen).

Voor de operationalisatie van de factoren naar concrete vragen (oftewel de items van een vragenlijst) is gebruik gemaakt van een groot aantal vragenlijsten/onderzoeksinstrumenten die zijn gebruikt voor onderzoek naar de hierboven beschreven modellen (onder andere Wanberg et al, 2002; Blau, 1994; Vinokur & Caplan, 1987; Schellekens, Langkamp & de Vries, 2005). Zie Schellekens et al. (2007) en Brouwer et al. (2011) voor een uitgebreide beschrijving van de instrumenten/schalen op basis waarvan de eerste conceptversie van de Werkverkenner is samengesteld. In een aantal gevallen betrof dit een vertaling van een Engelstalig (onderzoeks)instrument.

De op deze wijze tot stand gekomen vragenlijst bestond uit ongeveer 550 vragen. Vanwege de grootte van deze vragenlijst is deze in drie deelvragenlijsten opgesplitst, die in een cross-sectioneel onderzoek elk in een vergelijkbare steekproef zijn afgenomen. Hierbij is gekeken naar verschillen tussen personen met een WW-uitkering die langere tijd werkloos zijn en snelle werkhervatters. De laatste groep bestaat uit personen die binnen een jaar nadat zij werkloos zijn geworden weer een baan hebben gevonden. De resultaten van dit cross-sectionele onderzoek gaven inzicht in hoe goed de items onderscheid maakten tussen de groep langdurige werklozen en de groep snelle werkhervatters. Op basis van de analyses van de resultaten van dit cross-sectionele onderzoek zijn de drie deelvragenlijsten met in totaal 550 items gereduceerd tot één vragenlijst van 155 items. Deze 155 items corresponderen met 70 factoren die potentieel voorspellend zijn voor werkhervatting binnen één jaar. Ook na reductie van het aantal items zijn de drie belangrijkste theoretische modellen (het multidisciplinaire model van Wanberg, TPB en het motivatiemodel van Vroom) nog steeds vertegenwoordigd in de vragenlijst. Het deel van de vragenlijst waarin de “zachte” competenties en de persoonlijkheid werden uitgevraagd zijn weggelaten uit de vragenlijst, omdat deze geen onderscheid maakten tussen de groepen.

Voor de toetsing van de voorspellende waarde van factoren is een cross-sectioneel onderzoek niet voldoende. Dit kan alleen onderzocht worden in een longitudinaal onderzoek die de factoren van werkzoekenden in een vroeg stadium vastlegt en waarbij de werkzoekenden een bepaalde periode worden gevolgd. De gereduceerde vragenlijst uit het cross-sectionele onderzoek vormde daarom de basis voor een vervolg onderzoek dat longitudinaal van opzet was.

Het longitudinale onderzoek is uitgevoerd bij de WW-instroom in één regio in Nederland (Noord-Holland) in de jaren 2008 en 2009. Gedurende een jaar kregen WW-klienten die recent werkloos waren geworden de vragenlijst voorgelegd. Vervolgens werden zij een jaar gevolgd en werd bijgehouden of zij al dan niet het werk hervatten. Uit het onderzoek bleek dat tien van de ruim 70 factoren een voorspellende waarde hebben voor werkhervatting binnen twaalf maanden. De tien factoren samen voorspelden de status werkhervatting binnen een jaar vanaf aanvang van de WW-uitkering voor 69,2% van de werkzoekenden correct (Brouwer et al., 2015). Voor inzicht in de kans op werkhervatting volstaat dus een set van 10 harde en zachte factoren. Negen van deze factoren worden uitgevraagd met behulp van 18 items. Eén van de factoren (leeftijd) komt niet in de vragenlijst zelf voor, maar komt uit de UWV administratie. In overleg met het Werkbedrijf van UWV werd er nog één factor aan toegevoegd: lichamenlijk werkvermogen. Het resulterende instrument met 11 factoren (bestaande uit 20 items) werd gedoopt tot **Werkverkenner 1.0**. Deze versie wordt momenteel gebruikt op alle vestigingen van UWV voor diagnose en dienstverlening op maat. De theoretische modellen zijn alle drie vertegenwoordigd in Werkverkenner 1.0 (voor een overzicht van de factoren die in de online vragenlijst worden uitgevraagd, zie de factoren met een “ja” in het overzicht van tabel 1.1).

Tabel 1.1: Overzicht van de factoren die zijn opgenomen voor de doorontwikkeling van de Werkverkenner.  
 Voor een overzicht van de gestelde vragen per factor, zie Appendix A

	Factoren	Versie 1.0	Theoretisch model
<i>Opgenomen in vragenlijst op basis van aanvullende analyses op het onderzoekbestand van Noord-Holland</i>			
	<u>Harde factoren in vragenlijst</u>		
1	Aanvullende uitkeringen		Motivatie
2	Aanvullende inkomsten uit werk naast WW		Wanberg
3	Aantal jaren in laatste functie	ja	Wanberg
4	Fractie zelf ontslag genomen		Wanberg
5	Positie huishouden <sup>1</sup>		Wanberg
6	Problemen met de Nederlandse taal	ja	Wanberg
	<u>Zachte factoren in vragenlijst</u>		
7	Visie op terugkeer naar werk	ja	Wanberg
8	Werkzoekgedrag m.b.t direct contact werkgevers	ja	TPB
9	Werkzoekintentie	ja	TPB
10	Werkzoekattitude m.b.t voordeel en aangenaamheid		TPB
11	Werkzoekattitude m.b.t nut en noodzaak		TPB
12	Externe variabele attributie	ja	Motivatie
13	Algemeen werkvermogen	ja	Wanberg
14	Gezondheidsbeleving <sup>2</sup>	ja	Wanberg
15	Belemmering: financiële problemen		Wanberg
16	Belemmering: gebrek kinderopvang		Wanberg
17	Aantal uren in staat te werken		Wanberg
18	Balans voor- en nadelen van niet werken		Motivatie
19	Acceptatiebereidheid m.b.t. afwijkende tijden		Wanberg
20	Acceptatiebereidheid m.b.t. voltijd baan		Wanberg
21	Acceptatiebereidheid m.b.t. onwenselijk werk		Wanberg
22	Financiële noodzaak om snel werk te vinden		Wanberg
23	Subjectieve norm familie en partner		TPB
<i>Toegevoegd aan vragenlijst op basis van theoretische overwegingen (TPB-model)</i>			
24	Self-efficacy m.b.t. sollicitatie voorbereiding		TPB
25	Werkzoekgedrag m.b.t. sollicitaties versturen		TPB

	Factoren	Versie 1.0	Theoretisch model
<i>Harde factoren, toegevoegd uit administratie UWV op basis van analyses op het onderzoekbestand van Noord-Holland</i>			
26	Leeftijd	ja	Wanberg
27	Geslacht		Wanberg
28	Opleidingsniveau		Wanberg
29	Bedrijfstak		Wanberg
30	Bemiddelingsberoep (soort)		Wanberg
31	Bemiddelingsberoep (niveau)		Wanberg
32	Nationaliteit		Wanberg
33	Gemiddeld aantal arbeidsuren voor werkloosheid		Wanberg
34	Werkloosheidsomvang in uren		Wanberg
<i>Harde factoren, toegevoegd uit administratie UWV ten behoeve van het huidige onderzoek</i>			
35	Maximale duur WW-recht		Wanberg
36	WIA 35-min-indicatie <sup>3</sup>		Wanberg
37	Koplopervestiging		Wanberg

<sup>1</sup>In combinatie met de factor geslacht (nr. 27) wordt de factor Samenstelling huishouden gevormd en gebruikt in de analyses.

<sup>2</sup>Deze factor is berekend als de gemiddelde score op vier items: Psychisch werkvermogen, Lichamelijk werkvermogen, Ervaren gezondheid en Te ziek voelen om te werken.

<sup>3</sup>Een WW'er met een WIA 35-min-indicatie is afgewezen voor een WIA-uitkering (WIA: Wet werk en inkomen naar arbeidsvermogen). Deze mensen zijn eerst twee jaar ziek geweest, maar omdat hun inkomensverlies minder dan 35% is, hebben zij geen recht op een uitkering op grond van de WIA. Door de wijze van claimbeoordeling van de WIA kunnen sommige WW'ers met een WIA 35-min-indicatie toch ernstige beperkingen hebben die het vinden van een baan kunnen belemmeren.

#### 1.1.4 Op weg naar Werkverkenner 2.0

Het longitudinale onderzoek dat in de voorgaande paragraaf is besproken, was uitgevoerd in één regio van Nederland en er was geen onderscheid gemaakt tussen verschillende mogelijk relevante groepen binnen de groep werkzoekenden (bijv. 55 plussers). Verder was het onderzoek uitgevoerd in 2008 en 2009 en wilde UWV een actualisering van de Werkverkenner. Om die redenen is in 2014 het onderzoek naar de landelijke doorontwikkeling gestart. Ten behoeve van deze doorontwikkeling is een uitgebreide onderzoekversie van de Werkverkenner (1.0+) samengesteld, die een jaar lang op 11 zogenaamde koplopervestigingen van UWV is gebruikt (op welke vestigingen en in welke periode wordt beschreven in 3.1). Werkverkenner 1.0+ bestaat uit ongeveer 50 vragen, waaronder alle 20 reguliere vragen van Werkverkenner 1.0, aangevuld met 30 vragen uit het eerdere Noord-Holland onderzoek. Aanvullende analyses op het onderzoekbestand in Noord-Holland lieten zien dat bepaalde factoren een mogelijk voorspellende waarde kunnen hebben voor specifieke groepen binnen de WW. Deze vragen zijn toegevoegd. Verder zijn de extra harde factoren toegevoegd uit de administratie van UWV (o.a. geslacht en bedrijfstak). Werkverkenner 1.0+ wordt in het huidige onderzoek gebruikt (zie tabel 1.1) en zal resulteren in Werkverkenner 2.0.

## 1.2 Het huidige onderzoek

De vraagstellingen van het onderzoek naar de doorontwikkeling van de Werkverkenner zijn als volgt:

- 1) Wat zijn voorspellende persoons- en situatienmerken voor werkhervatting en duur van de werkloosheid van de WW-instroom?
- 2) Hoe kunnen de onderzoeksresultaten bij vraag 1 vertaald worden naar een objectief scoringssysteem (specificaties) voor een instrument dat direct toepasbaar is in de dienstverlening van UWV?
- 3) Wat zijn de psychometrische eigenschappen van het instrument?

Het onderzoek naar goede voorspelregels is van belang om te bepalen hoe het instrument in de toekomst nog steeds gebruikt kan worden als instrument om **dienstverlening op maat** te faciliteren. Op basis van de voorspelde kans op werkhervatting zullen werkzoekenden geselecteerd worden voor intensievere dienstverlening. Daarnaast wil UWV inzicht in de voorspelde werkloosheidsduur om een meer nauwkeurige inschatting te maken van de werkloosheidsduur bij aanvang van de WW en om de adviseur te ondersteunen bij het bepalen van de arbeidsmarktkansen van de werkzoekende. De uiteindelijke set van voorspellers zal worden vertaald naar Werkverkenner 2.0. De Werkverkenner 2.0 zal per werkzoekende inzicht geven in:

- > De kans op werkhervatting op specifieke momenten in de tijd na aanvang WW. Gegeven het feit dat men al tien weken WW heeft, doen we dit op 6, 9 en 12 maanden na instroom van de WW.
- > Welke factoren per werkzoekende belemmerend, dan wel bevorderend zijn. Daarmee dus ook een snelle diagnose van de werkzoekende (en de mogelijkheid tot inzet van dienstverlening die de kans kan doen vergroten).

Met werkhervatting in dit onderzoek bedoelen we werkhervatting waarbij de werkzoekende volledig uit de WW stroomt of al geen WW-uitkering meer heeft. Als men tijdens de WW-uitkering gedeeltelijk werkt, dan wordt de werkzoekende in dit onderzoek nog als werkloos beschouwd (zie box 1 voor de volledige definities gehanteerd in het onderzoek).

### **Box 1. Gehanteerde definities van werkloosheidsduur en werkhervatting in dit onderzoek:**

De werkloosheidsduur is de periode vanaf de aanvang van een WW-uitkering tot aan werkhervatting.

Er is sprake van een werkhervatting op het moment dat de werkzoekende het werk hervat binnen een bepaalde periode nadat de WW uitkering toegekend is en ook zijn WW-uitkering (al) volledig beëindigd is. Dat wil zeggen dat als een werkzoekende deels weer het werk hervat binnen een bepaalde periode, maar ook deels een WW-uitkering behoudt, dit niet beschouwd wordt als een werkhervatting, maar als werkloosheid.

De gehanteerde definitie van werkhervatting in dit onderzoek wijkt af van de definitie die gebruikt is in het voorafgaande onderzoek naar de voorspellers van werkhervatting in Noord-Holland. In het voorafgaande onderzoek werd werkhervatting vastgesteld als in het WW-systeem de uitkering binnen 12 maanden werd beëindigd wegens de reden werkhervatting. Wanneer de uitkering werd beëindigd wegens het bereiken van de maximale duur van de uitkering of wegens overige redenen, dan werden deze respondenten buiten de onderzoekspopulatie gelaten.

In dit onderzoek wordt een werkhervatting vastgesteld als in het WW-systeem de uitkering



beëindigd is wegens de reden werkhervatting of als de WW-uitkering beëindigd is wegens een andere reden en er sprake is van een dienstverband bij een werkgever in het systeem Polis. Deze gegevens worden door de belastingdienst geleverd aan UWV. Het voordeel van deze definitie is dat er een werkhervatting vastgesteld kan worden nadat de WW-uitkering beëindigd is wegens andere redenen dan werkhervatting. Dus: WW'ers waarbij de uitkering wordt beëindigd wegens andere redenen dan werkhervatting blijven onderdeel uitmaken van de onderzoekspopulatie. Het nadeel van het systeem Polis is dat hierin geen werkhervattingen als zelfstandig ondernemer worden geregistreerd. Dit is overigens maar een klein aandeel van het totaal aantal werkhervattingen. Het is wel mogelijk om werkhervattingen als zelfstandig ondernemer uit het WW-systeem te selecteren, zolang men recht op WW heeft

In deze eindrapportage richten we ons zowel op de eerste en derde onderzoeksvraag en doen we suggesties voor het stellen van een diagnose en de vormgeving van de feedback naar de adviseurs werk (de UWV professionals die met de Werkverkenner werken). In een apart document voor UWV zullen we het objectieve scoringsysteem beschrijven (onderzoeksvraag 2).

In deze eindrapportage beschrijven we allereerst de populatie en de steekproef van het onderzoek (H2) en gaan we vervolgens in op een gedeelte van de eerste onderzoeksvraag, namelijk: Wat zijn de voorspellende persoons- en situatiekenmerken voor werkhervatting binnen 6, 9 en 12 maanden voor de totale groep werkzoekenden? (H3). Vervolgens gaan we in op het tweede gedeelte van de eerste onderzoeksvraag, de voorspelling van werkloosheidsduur (H4). De representativiteit van de steekproef wordt beschreven in hoofdstuk 5. Hoofdstuk 6 gaat over het stellen van de diagnose (i.e. een verdere uitwerking van onderzoeksvraag 1): het bepalen welke mogelijkheden en belemmeringen een werkzoekende heeft om werk te vinden en geven we suggesties voor de vormgeving van de feedback naar de adviseurs werk. In hoofdstuk 7 worden de psychometrische eigenschappen van de Werkverkenner onderzocht (onderzoeksvraag 3). In het laatste hoofdstuk (H8) maken we onderscheid tussen verschillende klantgroepen (d.w.z. klantgroep 50-54, klantgroep 55+ en klantgroep WIA 35-min).

## 2 Beschrijving van de onderzoeksgroep

### 2.1 Beschrijving van de onderzoekspopulatie

De Werkverkenner werd tijdens de instroomperiode van het huidige onderzoek (van 1 maart 2014 t/m 28 februari 2015) ingevuld door WW'ers die gebruik maakten van de digitale werkmap van het UWV en langer dan drie maanden recht hebben op een WW-uitkering. De Werkverkenner werd destijds tussen de zesde en tiende week na toekenning van de WW-uitkering aangeboden. WW'ers hebben dan vervolgens twee weken de tijd om de Werkverkenner in te vullen. In de data zien we dat het aandeel werkzoekenden dat de Werkverkenner invult tot de tiende week van de werkloosheid nog toeneemt en vanaf de tiende week vrijwel stabiel is. Daarom zijn alleen die WW'ers geselecteerd voor de onderzoekspopulatie die én langer dan drie maanden recht hebben op WW én tien weken na aanvang nog een WW-uitkering hebben. WW'ers die geen gebruik maken van de digitale werkmap volgen een ander begeleidingstraject. De populatie voor dit onderzoek is gedefinieerd aan de hand van de volgende criteria:

- 1) werkzoekenden die recht hebben op meer dan drie maanden WW-uitkering;
- 2) werkzoekenden die de beschikking hebben over het e-kanaal;
- 3) werkzoekenden die op tien weken na aanvang van de WW nog steeds een WW-uitkering hebben (dus niet zijn uitgestroomd);
- 4) werkzoekenden die werkloos zijn geworden uit alle sectoren, behalve de sector overheid<sup>4</sup>;
- 5) werkzoekenden die binnen 12 weken gebruik maken van de werkmap en/of een actief digitaal CV hebben (gebruik digitale dienstverlening UWV);
- 6) werkzoekenden woonachtig in Nederland;
- 7) werkzoekenden waarbij de vaststelling van het WW-recht binnen tien weken bepaald is<sup>5</sup>.

Ten tijde van dit onderzoek kende UWV landelijk 50 vestigingen<sup>6</sup> waar WW'ers instroomden. We gebruiken in het huidige onderzoek gegevens van 11 vestigingen waar per 1 januari 2014 de Werkverkenner 1.0+ werd afgenomen. Deze 11 vestigingen zijn Almere, Amsterdam Oost, Arnhem, Breda, Dordrecht, Haarlem, Leeuwarden, Gouda, Zwolle, Sittard en Tilburg. Dit zijn zogenaamde koplopervestigingen. Deze vestigingen zijn aangewezen door UWV als vestigingen waarin innovatieve dienstverlening (o.a. de Werkverkenner) kan

---

<sup>4</sup> Bij de overheid is niet UWV, maar de voormalig werkgever verantwoordelijk voor de re-integratie.

Aangezien UWV geen taak heeft bij deze werkzoekenden, is gekozen om hen ook niet de Werkverkenner aan te bieden. De overheidssectoren omvatten ongeveer 4% van alle WW-instroom en ongeveer 16% van alle dienstverbanden. Het gaat om de sectoren: onderwijs en wetenschappen; rijk, politie en rechterlijke macht; defensie; provincies, gemeenten en waterschappen; openbare nutsbedrijven en overige overheidsinstellingen.

<sup>5</sup> De Werkverkenner wordt, ten tijde van dit onderzoek, aangeboden binnen zes en tien weken na toekenning WW-uitkering. Dus alleen werkzoekenden bij wie het WW-recht binnen tien weken is vastgesteld behoren tot de onderzoekspopulatie.

<sup>6</sup> Stand november 2014. Het aantal vestigingen is in de loop van de tijd afgenomen. Nu is er één vestiging per arbeidsmarktregio (dit zijn er 35).

worden uitgeprobeerd. Criteria voor een vestiging om als koploper in aanmerking te komen zijn: de vestiging wil meewerken aan innovatieve dienstverlening en de vestiging heeft zicht op continuïteit (i.e. wordt niet binnen korte termijn opgeheven). Het aandeel van de totale WW instroom dat de koplopervestigingen hebben is ruim een kwart. De ligging van de koploper-vestigingen is afgebeeld in figuur 2.1.

Wanneer we het in dit rapport hebben over de WW-instroom, dan gaat het altijd om de onderzoekspopulatie volgens de criteria zoals hierboven beschreven en zoals opgenomen in figuur 2.2. De populatie is hier dus niet de totale WW-populatie, maar betreft de selectie van werkzoekenden binnen de WW-instroom waar de Werkverkenner voor bestemd was gedurende de onderzoeksperiode.

## De koplopers



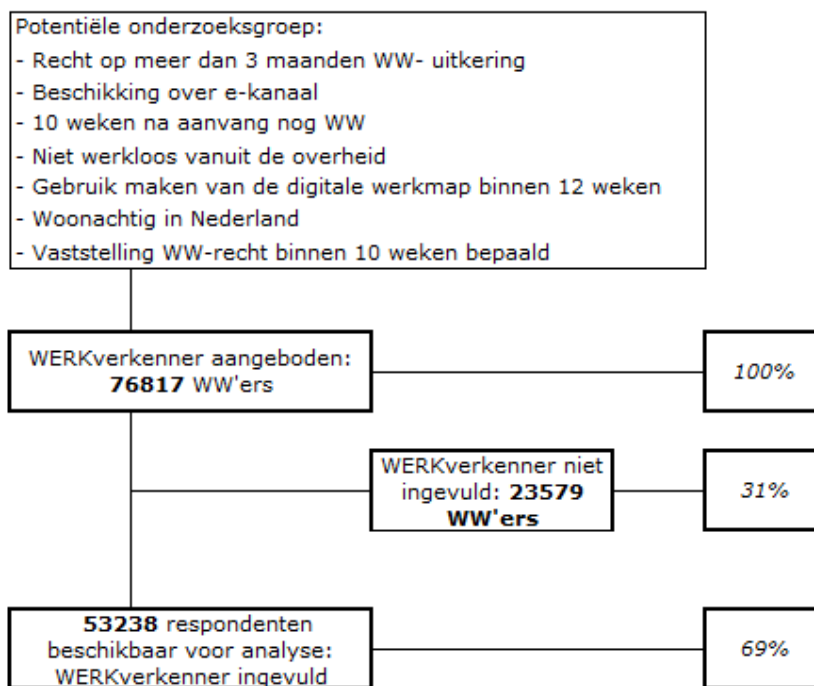
Figuur 2.1: Geografische spreiding van de 11 koplopervestigingen in Nederland.

Het eindbestand betreft de WW'ers die zijn ingestroomd in de WW van 1 maart 2014 tot en met 28 februari 2015. Tijdens deze periode is de manier van werving van de werkzoekenden gelijk gebleven. Uiteindelijk waren gegevens van 53.238 werkzoekenden beschikbaar voor de analyses. In figuur 2.2 is een stroomdiagram weergegeven van de aantallen in de steekproef van dit onderzoek.

## 2.2 Selectie van werkzoekenden uit steekproef

**Conclusie:** Van de 53.238 werkzoekenden die de Werkverkenner hebben ingevuld, zijn er 53.079 werkzoekenden waarvan we voor alle variabelen informatie hebben. Deze 53.079 werkzoekenden gebruiken we voor de analyses.

Voor de analyses selecteren we de werkzoekenden die géén missende waarden hebben op de variabelen die meegaan in de analyses. Tijdens de analyses worden werkzoekenden die een missende waarde op een (of meerdere) variabele(n) hebben namelijk buiten beschouwing gelaten. Er zijn 53.238 werkzoekenden beschikbaar voor de analyses (zie ook figuur 2.2). Van deze groep werkzoekenden is in tabel 2.1 een overzicht gegeven van het aandeel missende gegevens per verklarende variabele. Minder dan 0,30% van de beschikbare werkzoekenden heeft missende waarden op één of meerdere variabelen.



Figuur 2.2: Stroomdiagram van de steekproef werkzoekenden die gebruikt is voor het huidige onderzoek.

Er bestaan speciale methoden om werkzoekenden met missende waarden toch mee te kunnen nemen in de data-analyses, zoals bijvoorbeeld imputatietechnieken. Het aandeel werkzoekenden met missende waarden is hier echter zeer klein, waardoor we besluiten om geen speciale technieken toe te passen voor missende data. Na selectie van werkzoekenden met complete data zijn er 53.079 werkzoekenden die gebruikt worden voor de analyses.

## 2.3 Beschrijvende kenmerken van de steekproef

Met behulp van beschrijvende gegevens wordt inzicht gegeven in de kenmerken van de werkzoekenden die potentieel voorspellend zijn voor werkhervatting en de afhankelijke variabele (status werk bij 12 maanden) in het eindbestand. Deze kenmerken zijn uitgevraagd via de uitgebreide versie van de Werkverkenner (zie bijlage A) en relevante kenmerken die

UWV bijhoudt in haar administratie (zie tabel 1.1). We splitsen de beschrijving van de kenmerken uit naar werkhervattingstatus. Op deze manier kunnen we per kenmerk bekijken of de groep die binnen 12 maanden het werk heeft hervat verschilt van de groep die het werk nog niet heeft hervat op 12 maanden.

Voor het berekenen van de beschrijvende statistieken maken we onderscheid tussen categorische en continue kenmerken. Categorische kenmerken zijn discrete kenmerken, zoals bijvoorbeeld man of vrouw voor het kenmerk geslacht. Bij categorische kenmerken is de volgorde van de categorieën willekeurig. Ook de “afstanden” tussen de categorieën hebben geen betekenis (bijv. de afstand tussen categorie 1 en 2 hoeft niet hetzelfde te zijn als tussen categorie 2 en 3; hierdoor kan er geen gemiddelde berekend worden). Voor deze kenmerken rapporteren we percentages; deze staan weergegeven in tabel 2.1. Tevens rapporteren we de effectgrootte (Cramer's *V*) voor het verschil tussen de twee groepen. Het gaat hierbij om de groep werkhervatters binnen 12 maanden en de groep respondenten die nog steeds werkzoekend zijn op 12 maanden. Cramer's *V* kan waarden aannemen tussen 0 en 1, waarbij de waarde 0,1 in de literatuur wordt beschouwd als een *klein* effect, 0,3 als een *middelgroot* effect en 0,5 als een *groot* effect (Cohen, 1992). In dit onderzoek beschouwen we daarom een (afgeronde) waarde van Cramer's *V* kleiner dan 0,1 als *zeer klein*, een waarde tussen 0,1 en 0,3 als *klein*, tussen 0,3 en 0,5 als *middelgroot* en 0,5 of hoger als *groot*.

Tabel 2.1 laat zien dat de meeste effectgroottes klein zijn, met uitzondering van de factoren Aantal jaren in laatste functie (Cramer's *V* = 0,27), Leeftijd in klassen (Cramer's *V* = 0,34) en Maximale duur WW-recht (Cramer's *V* = 0,26); deze laatste drie factoren laten (afgerond) een middelgroot effect zien. Dit betekent dat, over het algemeen genomen, de verdelingen van de percentages per factor weinig verschillen tussen de twee groepen.

Tabel 2.1: Beschrijvende statistieken (%) van de categorische kenmerken van de werkzoekenden in het eindbestand (*N* = 53.079).

WV 1.0	Categorische variabelen	Totaal	Geen werk binnen 12 maanden ( <i>n</i> = 25.409)	Werk hervat binnen 12 maanden ( <i>n</i> = 27.670)	Verskil tussen de twee groepen (Cramer's <i>V</i> )	Missende waarden <sup>2</sup> <i>n</i> (%)
	Aanvullende uitkering(en) naast WW (% ja)				0,10 (K)	0
	<i>Nee</i>	92,3	89,5	94,9		
	<i>Ja</i>	7,7	10,5	5,1		
	Aanvullende inkomsten uit werk naast WW				0,03 (Z)	0
	<i>Nee</i>	85,6	86,7	84,6		
	<i>Ja</i>	14,4	13,3	15,4		
X	Aantal jaren in laatste functie				0,27 (M)	
	<i>0 – 0,5 jaar</i>	12,5	8,4	16,3		
	<i>meer dan 0,5 - 1 jaar</i>	12,0	8,7	14,9		
	<i>meer dan 1 - 2 jaar</i>	14,7	12,4	16,9		
	<i>meer dan 2 - 5 jaar</i>	23,6	20,8	26,2		
	<i>meer dan 5 - 10 jaar</i>	15,6	18,5	13,0		
	<i>meer dan 10 - 50 jaar</i>	21,5	31,1	12,7		
	Samenstelling huishouden				0,18 (K)	0
	<i>alleenstaande ouder</i>	7,1	8,7	5,5		

WV 1.0	Categorische variabelen	Totaal	Geen werk binnen 12 maanden (n = 25.409)	Werk hervat binnen 12 maanden (n = 27.670)	Vershil tussen de twee groepen (Cramer's V)	Missende waarden <sup>2</sup> n (%)
	<i>man getrouwd/woont samen, jongste kind thuis 0-6 jaar</i>	7,4	5,0	9,5		
	<i>vrouw getrouwd/woont samen, jongste kind thuis 0-6 jaar</i>	10,1	10,2	9,9		
	<i>getrouwd/woont samen, jongste kind thuis 7 jaar en ouder</i>	22,9	24,1	21,7		
	<i>getrouwd/woont samen, geen thuiswonende kinderen</i>	25,9	29,7	22,4		
	<i>woont alleen</i>	18,4	17,6	19,2		
	<i>anders</i>	8,3	4,7	11,7		
X	Problemen Nederlandse Taal				0,02 (Z)	0
	<i>geen problemen</i>	94,3	94,2	94,4		
	<i>beetje problemen</i>	4,1	4,3	3,9		
	<i>veel problemen/kan geen NL</i>	1,7	1,6	1,8		
	Balans voor- en nadelen niet-werken				0,14 (K)	0
	<i>meer voordelen dan nadelen</i>	1,8	2,3	1,4		
	<i>evenveel nadelen als voordelen</i>	23,1	29,1	17,6		
	<i>meer nadelen dan voordelen</i>	75,1	68,6	81,1		
	Financiële noodzaak om <2 mnd. werk te vinden				0,19 (K)	0
	<i>niet belangrijk</i>	10,8	15,2	6,7		
	<i>belangrijk</i>	51,7	55,5	48,2		
	<i>zeer belangrijk</i>	37,5	29,2	45,0		
X	Leeftijd in klassen				0,34 (M)	
	<i>jonger dan 28 jaar</i>	13,0	5,0	20,4		
	<i>tussen 28 en 39 jaar</i>	29,4	23,6	34,8		
	<i>tussen 40 en 49 jaar</i>	26,8	27,6	26,1		
	<i>tussen 50 en 54 jaar</i>	12,8	15,7	10,1		
	<i>55 jaar of ouder</i>	18,0	28,2	8,6		
	Geslacht				0,09 (K)	0
	<i>Man</i>	49,1	44,2	53,6		
	<i>Vrouw</i>	50,9	55,8	46,4		
	Opleidingsniveau				0,06 (K)	112 (0,0)
	<i>Laag</i>	20,0	22,3	17,9		
	<i>Middelbaar</i>	51,4	49,0	53,5		
	<i>Hoog</i>	28,6	28,6	28,6		
	Bedrijfstak voor WW				0,16 (K)	0
	<i>Landbouw, Visserij, Voeding</i>	3,5	3,2	3,8		
	<i>Bouw en Hout</i>	5,5	4,2	6,6		
	<i>Industrie sec</i>	10,3	11,0	9,7		
	<i>Winkelbedrijf en Groothandel</i>	16,5	17,3	15,7		
	<i>Transport</i>	5,7	5,3	6,0		
	<i>(Financ.) Dienstverlening</i>	21,0	23,3	18,9		
	<i>Uitzend</i>	10,0	6,0	13,6		

WV 1.0	Categorische variabelen	Totaal	Geen werk binnen 12 maanden (n = 25.409)	Werk hervat binnen 12 maanden (n = 27.670)	Verskil tussen de twee groepen (Cramer's V)	Missende waarden <sup>2</sup> n (%)
	<i>Gezondheid</i>	19,0	21,9	16,4		
	<i>Overig bedrijf en beroep</i>	8,5	7,7	9,2		
	<i>Overig</i>	0,1	0,1	0,2		
	Bemiddelingsberoep soort				0,13 (K)	80 (0,0)
	<i>Economisch-administratieve beroepen</i>	35,0	39,0	31,3		
	<i>Agrarische beroepen</i>	1,8	1,2	2,3		
	<i>Informatica beroepen</i>	2,4	2,4	2,4		
	<i>Medische en paramedische beroepen</i>	3,3	3,4	3,2		
	<i>Openbare orde- en veiligheidsberoepen</i>	,9	,8	1,0		
	<i>Pedagogische beroepen</i>	2,1	2,2	2,1		
	<i>Sociaal-culturele beroepen</i>	7,8	8,4	7,2		
	<i>Technische en industrieberoepen</i>	20,1	15,8	24,1		
	<i>Transportberoepen</i>	7,4	6,6	8,2		
	<i>Verzorgende en dienstverlenende beroepen</i>	19,1	20,1	18,2		
	Bemiddelingsberoep niveau				0,03 (Z)	80 (0,0)
	<i>Elementaire</i>	8,3	8,5	8,1		
	<i>Lagere</i>	31,0	30,0	31,9		
	<i>Middelbare</i>	36,9	36,7	37,2		
	<i>Hogere</i>	20,2	21,0	19,4		
	<i>Wetenschappelijk</i>	3,5	3,7	3,4		
	Nationaliteit				0,05 (K)	
	<i>1e nat: NL, 2e nat: geen of Westerse</i>	95,7	96,2	95,3		43 (0,0)
	<i>1e of 2e nat: Poolse</i>	1,3	,8	1,8		
	<i>1e of 2e nat: Niet-Westers (vooral Turks, Marokkaans)</i>	1,1	1,2	1,1		
	<i>1e nat: Westers, anders dan NL en Poolse</i>	1,8	1,9	1,8		
	Gemiddeld aantal arbeidsuren voor werkloosheid				0,09 (K)	
	<i>&lt; 12 uren per week</i>	2,1	2,2	1,9		
	<i>tussen 12 en 24 uren per week</i>	20,5	23,5	17,8		
	<i>tussen 25 en 32 uren per week</i>	20,6	21,7	19,6		
	<i>meer dan 32 uren per week</i>	56,8	52,6	60,6		
	Maximale duur WW-recht meer dan 12 mnd maanden				0,26 (M)	0
	<i>Nee</i>	24,3	12,6	35,0		
	<i>Ja</i>	75,7	87,4	65,0		
	WIA 35-min-indicatie				0,09 (K)	0
	<i>Nee</i>	98,0	96,6	99,2		
	<i>Ja</i>	2,0	3,4	0,8		
	Koplopervestiging (district)				0,07 (K)	0
	<i>Almere (district: Utrecht en Flevoland)</i>	7,3	7,4	7,3		
	<i>Amsterdam Oost (district: Groot Amsterdam)</i>	7,6	8,2	7,1		
	<i>Arnhem (district: Gelderland Midden en Zuid)</i>	9,4	9,9	8,9		
	<i>Breda (district: West Brabant en Zeeland)</i>	15,7	16,6	14,7		

WV 1.0	Categorische variabelen	Totaal	Geen werk binnen 12 maanden (n = 25.409)	Werk hervat binnen 12 maanden (n = 27.670)	Verskil tussen de twee groepen (Cramer's V)	Missende waarden <sup>2</sup> n (%)
	<i>Dordrecht (district: Rijnmond)</i>	7,2	7,5	6,9		
	<i>Haarlem (district Noord Holland Noord)</i>	7,8	7,8	7,7		
	<i>Midden Holland (district: Den Haag en Leiden)</i>	15,6	13,7	17,4		
	<i>Leeuwarden (district: Noord)</i>	3,4	3,6	3,3		
	<i>Zwolle (district: Overijssel en Gelderland Noord)</i>	10,9	9,6	12,0		
	<i>Sittard (district: Limburg)</i>	4,9	5,1	4,8		
	<i>Tilburg (district: Midden- en Oost-Brabant)</i>	10,2	10,6	9,8		

Noot. De aangekruiste items in de eerste kolom zijn onderdeel van Werkverkenner 1.0 (WV 1.0); Z = zeer klein effect (Cramer's  $V < 0,1$ ); K = klein effect ( $0,1 \leq V < 0,3$ ); M = middelgroot effect ( $0,3 \leq V < 0,5$ ).



Tabel 2.2: Beschrijvende statistieken van continue factoren in het eindbestand (N = 53.079)

WV 1.0	Continue variabelen	Totale groep		Geen werk binnen 12 maanden (n = 25.409)					Werkhervatting binnen 12 maanden (n = 27.670)					Verskil tussen de twee groepen:	Missende waarden in steekproef <sup>2</sup>
		M	SD	M	SD	Med	min	max	M	SD	Med	min	max	Cohen's d	n (%)
X	Aantal jaren in laatste functie	6,7	8,5	8,9	9,8	5,0	0,0	48,1	4,7	6,6	2,4	0,0	48,0	0,50 (M)	0
	Fractie zelf ontslag genomen	0,1	0,3	0,1	0,3	0,0	0,0	1,0	0,2	0,4	0,0	0,0	1,0	0,21 (K)	1 (0,0)
X	Visie op terugkeer naar werk [S]	3,5	0,8	3,3	0,8	3,3	1,0	5,0	3,7	0,7	3,7	1,0	5,0	0,56 (M)	0
X	Werkzoekgedrag m.b.t. direct contact werkgevers [S]	1,9	0,6	1,8	0,6	1,7	1,0	4,0	2,0	0,6	2,0	1,0	4,0	0,42 (M)	0
X	Werkzoekintentie [S]	4,1	0,7	4,1	0,7	4,0	1,0	5,0	4,2	0,7	4,3	1,0	5,0	0,26 (K)	0
	Werkzoekattitude m.b.t. voordeel en aangenaamheid [S]	3,3	0,8	3,3	0,8	3,0	1,0	5,0	3,3	0,8	3,0	1,0	5,0	0,04 (Z)	0
	Werkzoekattitude m.b.t. nut en noodzaak [S]	4,3	0,7	4,2	0,7	4,0	1,0	5,0	4,4	0,6	4,5	1,0	5,0	0,30 (K)	0
X	Externe variabele attributie [S]	3,3	0,7	3,4	0,7	3,3	1,0	5,0	3,3	0,7	3,3	1,0	5,0	0,14 (Z)	0
X	Algemeen werkvermogen	8,0	1,7	7,6	1,8	8,0	1,0	10,0	8,3	1,4	8,0	1,0	10,0	0,48 (M) <sup>5</sup>	0
X	Gezondheidsbeleving <sup>3</sup> [S]	4,1	0,7	3,9	0,8	4,0	1,0	5,0	4,2	0,6	4,3	1,0	5,0	0,50 (M)	0
	Belemmering: problemen financiële situatie	1,9	1,1	2,0	1,1	2,0	1,0	5,0	1,9	1,1	1,0	1,0	5,0	0,16 (K) <sup>5</sup>	0
	Belemmering: problemen kinderopvang <sup>4</sup>	1,3	0,8	1,4	0,8	1,0	1,0	5,0	1,3	0,8	1,0	1,0	5,0	0,05 (Z)	0
	Belemmering: aantal uren in staat om te werken	35,1	11,3	32,1	11,9	32,0	0,0	50,0	37,7	10,0	40,0	0,0	50,0	0,51 (M)	0
	Acceptatiebereidheid m.b.t. afwijkende tijden [S]	2,7	1,2	2,5	1,2	2,5	1,0	5,0	2,9	1,2	3,0	1,0	5,0	0,34 (K)	0
	Acceptatiebereidheid voltijd baan [S]	3,6	1,5	3,2	1,6	4,0	1,0	5,0	3,9	1,4	5,0	1,0	5,0	0,44 (K)	0
	Acceptatiebereidheid m.b.t. onwenselijk werk [S]	2,9	0,9	2,8	0,9	3,0	1,0	5,0	2,9	0,9	3,0	1,0	5,0	0,06 (Z)	0
	Subjectieve norm familie en partner	9,7	4,2	9,0	4,2	9,0	0,0	16,0	10,3	4,2	9,0	0,0	16,0	0,31 (K)	0
	Self-efficacy m.b.t. sollicitatievoorbereiding [S]	3,9	0,6	3,9	0,6	4,0	1,0	5,0	3,9	0,6	4,0	1,0	5,0	0,08 (Z)	0
	Werkzoekgedrag m.b.t. sollicitaties versturen [S]	2,8	0,9	2,8	0,9	3,0	1,0	4,0	2,9	0,9	3,0	1,0	4,0	0,17 (Z)	0
X	Leeftijd	42,1	11,5	46,3	10,8	47,0	19,0	63,0	38,3	10,8	37,0	18,0	63,0	0,75 (G)	0
	Gemiddeld aantal arbeidsuren voor werkloosheid	32,4	8,9	31,6	9,0	34,0	1,0	77,0	33,2	8,8	36,0	1,0	80,0	0,17 (K)	3 (0,0)
	Werkloosheidsomvang in uren	31,0	9,8	30,2	9,9	32,0	1,0	77,0	31,8	9,7	36,0	2,0	80,0	0,16 (K) <sup>5</sup>	0
	Maximale duur van WW-recht in weken	97,0	48,1	113,2	45,2	121,0	13,0	219,0	82,1	45,7	73,0	13,0	217,0	0,68 (M)	0

1 De eerste kolom geeft met een kruis de items weer die onderdeel zijn van Werkverkenner 1.0.

2 Gebaseerd op 53.238 werkzoekenden; de gepresenteerde aantallen met missende waarden zijn de werkzoekenden die uiteindelijk zijn geëxcludeerd voor de analyses.

3 In Werkverkenner 1.0 zijn drie van de vier items behorende bij de factor Gezondheidsbeleving afzonderlijk meegenomen: Te ziek voelen om te werken, lichamelijk werkvermogen en geestelijk werkvermogen. Ervaren gezondheid is destijds niet meegenomen.

4 Aan huishoudens zonder thuiswonende kinderen is deze vraag niet gesteld. Deze huishoudens krijgen als antwoord 'geen problemen'.

5 We ronden Cohen's d vervolgens af op 1 decimaal en passen dan de indeling toe. Daarom wordt bijv. 0,48 als een middelgroot effect beschouwd en 0,16 als een klein effect.

M = gemiddelde, SD = standaard deviatie. Med = mediaan, min = minimum, max = maximum, S = schaal (gecreëerd uit meerdere items); Z = zeer klein/verwaarloosbaar ( $d < 0,2$ ); K = klein effect ( $0,2 \leq d < 0,5$ ); M = middelgroot effect ( $0,5 \leq d < 0,8$ ); G = groot effect ( $d \geq 0,8$ ).

Voor continue kenmerken geldt in principe dat deze, binnen een bepaald bereik, elke mogelijke tussenliggende waarde kunnen aannemen, zoals bijvoorbeeld Leeftijd in jaren (tussen 0 en 63 jaar is elke waarde mogelijk). Bovendien geldt voor continue kenmerken dat de afstand tussen twee waardes even groot is (bijv. de afstand tussen 1 en 2 is gelijk aan de afstand tussen 2 en 3). Voor deze kenmerken rapporteren we naast het gemiddelde ook de mediaan. De mediaan is het middelste getal wanneer alle waarden worden geordend van klein naar groot. Vijftig procent van de waarden ligt onder de mediaan en 50% van de waarden ligt boven de mediaan. Tevens rapporteren we de effectgrootte (Cohen's  $d$ ) voor het verschil tussen de twee groepen. Een Cohen's  $d$  van  $[0,20]$  wordt in de literatuur beschouwd als een *klein* effect,  $[0,50]$  als een *middelgroot* effect en  $[0,80]$  als een *groot* effect (Cohen, 1992). In dit onderzoek beschouwen we daarom een waarde van  $d$  kleiner dan  $[0,2]$  als zeer klein (oftewel verwaarloosbaar), een waarde tussen  $[0,2]$  en  $[0,5]$  als klein, tussen  $[0,5]$  en  $[0,8]$  als middelgroot en waarden gelijk aan  $[0,8]$  of hoger als groot. We ronden hierbij eerst af op 1 decimaal.

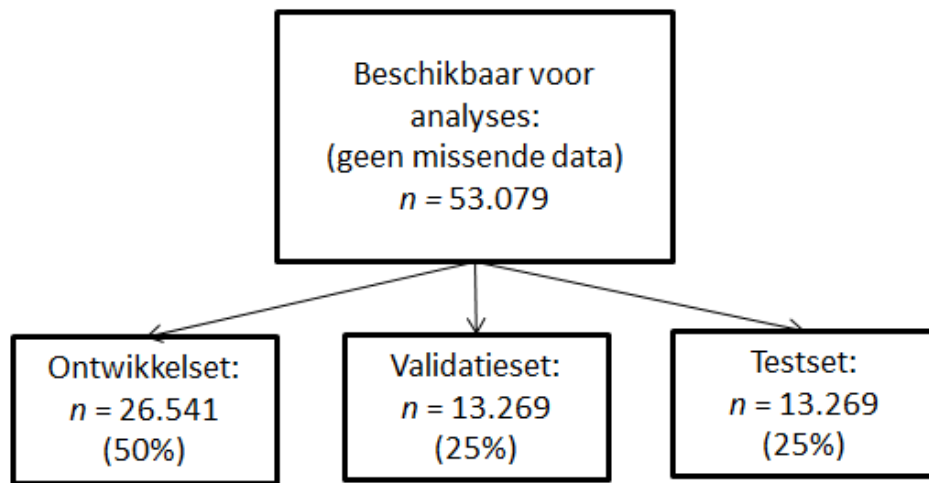
De continue kenmerken staan weergegeven in tabel 2.2. Voor een paar factoren hebben we zowel een categorische als continue variant opgenomen<sup>7</sup>. Het gaat om de volgende factoren: Aantal jaren in laatste functie, Leeftijd, Gemiddeld aantal arbeidsuren voor werkloosheid en Maximale duur recht op WW. Deze variabelen komen in beide tabellen voor. Tabel 2.2 laat zien dat de meeste effectgroottes zeer klein tot klein zijn (15 van de 23); Zeven effecten zijn middelgroot. De factor Leeftijd laat (afgerond) een groot effect zien ( $d = 0,75$ ). Dit betekent dat het grootste verschil is waargenomen (i.e. 0,75 standaarddeviatie) tussen de gemiddelde leeftijd van de groep werkhervatters (38,3 jaar) en de gemiddelde leeftijd van de groep respondenten die nog steeds werkzoekend zijn op 12 maanden (46,3 jaar). De statistische toetsen van de verschillen tussen de groepen bespreken we in het volgende hoofdstuk.

Voor de analyses verdelen we de dataset in een ontwikkelset (50% van de WW'ers in het bestand), een testset (25% van de WW'ers in het bestand) en een validatieset (25% van de WW'ers in het bestand)<sup>8</sup>. De ontwikkelset gebruiken we om de modellen te schatten. De validatieset gebruiken we om de beste modelspecificaties te kiezen (bijv. stapsgewijze selectie met het BIC criterium of met het AIC). Met de testset bepalen we vervolgens de accuraatheid van de predictie van de modellen voor toekomstige werkzoekenden, de zogenoemde gegeneraliseerde predictie accuraatheid. De werkzoekenden zijn aselekt (dat wil zeggen, op basis van toeval) toegewezen aan de drie sets. Een overzicht van de beschikbare werkzoekenden en de verdeling van de werkzoekenden over de drie sets is weergegeven in figuur 2.3. In hoofdstuk 3 zullen we zien dat de verdeling van status werkhervatting over de drie sets gelijk is (zie tabel 3.1).

---

<sup>7</sup> De twee redenen voor het opnemen van een categorische en continue variant zijn: 1) om niet-lineaire verbanden te kunnen onderzoeken (o.a., voor de factor Leeftijd) en 2) om multicollineariteit tegen te gaan (o.a. tussen de factoren Maximale duur recht op WW en Leeftijd). Dit komt verder aan de orde in hoofdstuk 3.

<sup>8</sup> Een andere vorm van toetsing is externe validatie, waarbij we met behulp van een nieuwe dataset het ontwikkelde model toetsen. Voor deze nieuwe dataset zouden we dan werkzoekenden gebruiken die zijn ingestroomd na februari 2015 (i.e. deze werkzoekenden vallen buiten het jaarcohort). Dit valt niet onder de scope van dit onderzoek, maar maakt onderdeel uit van het kwaliteitsprogramma periodiek onderhoud en doorontwikkeling Werkverkenner van UWV.



Figuur 2.3: Verdeling van de beschikbare werkzoekenden over de ontwikkel-, validatie- en testset.

## 3 Resultaten voorspelling werkhervatting

### 3.1 Onderzoeksvraag 1 en deelvragen

In dit hoofdstuk gaan we in op een gedeelte van onderzoeksvraag 1. We onderzoeken de invloed van persoons- en situatiekenmerken (i.e. de harde en zachte factoren weergegeven in tabel 1.1) op de status van werkhervatting binnen 12, 9 en 6 maanden na aanvang van de WW. We richten ons in de beschrijving eerst op de voorspelling voor 12 maanden, omdat deze het meest belangrijk is voor dit onderzoek; deze voorspelling kan namelijk vergeleken worden met de voorspelling van Werkverkenner 1.0. De resultaten voor de voorspelling voor 9 en 6 maanden worden daarna vermeld.

We beantwoorden de volgende deelvragen:

- 1a) Hoe voorspellend is iedere factor apart?
- 1b) Hoe voorspellend zijn de factoren samen?
- 1c) Wat is het meest eenvoudige model (d.w.z. een begrijpelijk model met zo min mogelijk factoren) waarmee we (bijna) even goed kunnen voorspellen als het model met alle factoren?

### 3.2 Status werkhervatting

Tabel 3.1 geeft de geobserveerde verdeling weer voor de totale dataset en de ontwikkel-, validatie- en testset van werkzoekenden die het werk hebben hervat binnen respectievelijk 12, 9 en 6 maanden. De verdelingen in het totale bestand komen overeen met de cijfers in de landelijke populatie. Zoals ook aangegeven in hoofdstuk 2 hebben we het hier niet over de *totale populatie* WW-instromers, maar over een selectie binnen de WW-instroom waar de Werkverkenner voor bedoeld is (zie figuur 2.2). De tabel laat zien dat de percentages werkhervatters in de ontwikkel-, validatie- en testset overeenkomen met het percentage in het totale bestand. De tabel laat verder zien dat 52% van de WW-instromers het werk binnen 12 maanden hervat, binnen 9 maanden heeft 43% het werk hervat en binnen 6 maanden 30%.

Tabel 3.1: Aantallen en percentages van de werkzoekenden die het werk hebben hervat binnen 12, 9 en 6 maanden<sup>1</sup>, verdeeld over de ontwikkel-, validatie- en testset.

	Werk hervat binnen 12 maanden <i>n</i> (%)	Werk hervat binnen 9 maanden <i>n</i> (%)	Werk hervat binnen 6 maanden <i>n</i> (%)	Totaal <i>n</i> (%)
<b>Totale bestand</b>	27670 (52%)	23035 (43%)	15986 (30%)	53.079 (100%)
<b>Ontwikkelset</b>	13857 (52%)	11564 (44%)	7998 (30%)	26.541 (100%)
<b>Validatieset</b>	6945 (52%)	5764 (43%)	3983 (30%)	13.269 (100%)
<b>Testset</b>	6868 (52%)	5707 (43%)	4005 (30%)	13.269 (100%)

<sup>1</sup> Het betreft het hervatten van werk in de periode vanaf tien weken t/m 6, 9 of 12 maanden na instroom in de WW. De aantallen zijn gebaseerd op de werkzoekenden die géén missende waarden hebben op de gegevens die meegaan in het predictiemodel.

### 3.3 Naïeve voorspelmethode

Allereerst kijken we naar de resultaten van de naïeve voorspelmethode. Volgens deze methode krijgt iedere respondent dezelfde voorspelling. De methode maakt geen gebruik van de invloed van voorspellende factoren in het model. Dat wil zeggen dat de scores van een werkzoekende op de factoren van de Werkverkenner niet gebruikt worden om de status werkhervatting te voorspellen. In plaats daarvan is de voorspelling gebaseerd op de meest voorkomende categorie van de uitkomstmaat (i.e. status werkhervatting), dus de categorie die het hoogste percentage heeft. Dit heeft als consequentie dat ofwel de specificiteit van de voorspelregel 100% is en de sensitiviteit 0%, ofwel de sensitiviteit 100% is en de specificiteit 0%. Bijvoorbeeld: wanneer we iedere werkzoekende bij aanvang van de WW de status werk hervat binnen 12 maanden zouden geven, dan voorspellen we  $6.868 / 13.269 * 100 = 52\%$  van de werkzoekenden correct (we gebruiken hiervoor de testset). Dit noemen we ook wel de *a priori* kans op werk hervatting binnen 12 maanden. De sensitiviteit van deze voorspelregel is gelijk aan 100% (alle daadwerkelijke werkhervatters worden correct voorspeld) en de specificiteit is 0% (alle niet-werkhervatters worden fout voorspeld).

Voor de voorspelling op 9 maanden zouden we volgens deze methode iedereen de status nog steeds werkzoekend op 9 maanden geven en daarmee **57%** correct voorspellen. Op 6 maanden geven we volgens deze methode iedereen de status nog steeds werkzoekend op 6 maanden en voorspellen daarmee **70%** correct. Het is nu interessant om te kijken of we een groter deel van de werkzoekenden correct kunnen classificeren met behulp van analysemodellen waarbij we wel de scores op de factoren van de Werkverkenner gebruiken.

### 3.4 Invloed van iedere factor apart op status werkhervatting

**Conclusie:** Op basis van de univariate analyses laten de volgende tien factoren de sterkste relatie zien met status werkhervatting binnen 12 maanden: Leeftijd, Maximale duur WW-recht (continue en categorische variant), Visie op terugkeer naar werk, Aantal jaren in laatste functie, Aantal uren in staat te werken, Gezondheidsbeleving, Algemeen werkvermogen, Acceptatiebereidheid m.b.t. een voltijdbaan en Werkzoekgedrag m.b.t. direct contact werkgevers. Ook voor status werkhervatting binnen 9 maanden en binnen 6 maanden laten dezelfde tien factoren de sterkste relatie zien. De volgende factoren verkleinen de kans op werkhervatting: Leeftijd, Maximale duur WW-recht en Aantal jaren in laatste functie (bijv. hoe langer de maximale duur van het recht op een WW uitkering hoe kleiner de kans op werkhervatting). De overige bovengenoemde factoren vergroten de kans op werkhervatting (bijvoorbeeld hoe hoger de score op visie op terugkeer naar werk, hoe groter de kans op werkhervatting). De univariate analyses laten voor alle factoren zien dat de aard van het verband in de verwachte richting is (op basis van de achterliggende theorie).

Met behulp van enkelvoudige logistische regressie analyses is per factor (uit tabel 1.1) bepaald hoe sterk de invloed was op status werkhervatting binnen 12, 9 en 6 maanden (deelvraag 1a). Hierbij was status werkhervatting binnen 12 maanden (of 9 of 6 maanden) de uitkomstvariabele en de factor de *voorspeller*. Deze zogenaamde univariate analyses zijn uitgevoerd op de ontwikkelset. Deze analyses hebben als doel te bepalen: 1) hoe sterk de invloed van de factor is en of de aard van het verband in de verwachte richting is (op basis van de achterliggende theorie); met andere woorden, te bekijken of met een hogere score op de factor de kans op werkhervatting naar verwachting toeneemt (of afneemt) en 2) of, voor een aantal variabelen, de categorische variant of de continue variant een sterker verband laat zien.

Voor het realiseren van het tweede doel, beperken we ons tot de variabelen Leeftijd, Maximale duur recht op WW, Aantal jaren in laatste functie en Gemiddeld aantal arbeidsuren voor werkloosheid<sup>9</sup>. We bepalen welke variant, de categorische of de continue, resulteert in de hoogste waarde van de Nagelkerke  $R^2$ ; deze maat geeft aan hoe goed het model past bij de data en is vergelijkbaar met de verklaarde variantie in lineaire regressie: hoe hoger de  $R^2$  des te beter de fit van het model. Ook kijken we naar de AIC; deze maat geeft aan hoe goed het model past bij de data rekening houdend met het aantal termen in het model. We vergelijken de waarde van de AIC van het model met de categorische variant met de waarde van de AIC van het model met de continue variant. Hoe lager de AIC, des te beter het model past. De absolute waarde van de AIC heeft geen betekenis in termen van goed of matig. Beide maten, de  $R^2$  en de AIC, worden ook wel goodness-of-fit maten genoemd.

Voor de variabele Aantal jaren in laatste functie resulteerde de categorische variant in een hogere Nagelkerke  $R^2$  dan de continue variant op alle meetmomenten. Tabel 3.2 geeft de resultaten voor status werkhervatting binnen 12 maanden weer. Voor de andere variabelen, Leeftijd, Maximale duur recht op WW en Gemiddeld aantal arbeidsuren voor werkloosheid, bleek de continue variant beter voor zowel 12, 9 als 6 maanden.

Tabel 3.2: Verschillen in goodness-of-fit maten van de categorische variant en de continue variant van vier factoren van de Werkverkenner. De resultaten betreffen univariate analyses op de ontwikkelset voor status werkhervatting binnen 12 maanden ( $n = 26.541$ ).

	Nagelkerke $R^2$		Akaike Information Criterion (AIC) <sup>1</sup>	
	Categorische variant	Continue variant	Categorische variant	Continue variant
Aantal jaren in laatste functie	0,091	0,080	34875	35094
Leeftijd	0,152	0,157	33548	33423
Maximale duur WW-recht	0,090	0,132	34901	33986
Gmd. aantal uren voor werkloosheid	0,010	0,011	36541	36518

<sup>1</sup> Voor Nagelkerke  $R^2$  geldt: een hogere waarde is een betere fit; voor AIC geldt: een lagere waarde is een betere fit (de absolute waarde van de AIC wordt niet geïnterpreteerd).

Voor de univariate analyses rapporteren we daarom voor Aantal jaren in laatste functie de categorische variant en voor de overige drie variabelen de continue. Op basis van de resultaten van de multivariate analyses (zie sectie 3.5) vinden we echter dat voor Maximale duur recht op WW en Gemiddeld aantal arbeidsuren voor werkloosheid de categorische variant te verkiezen is boven de continue om andere redenen, namelijk, multicollineariteit en niet-lineariteit. We geven daarom ter volledigheid voor deze twee variabelen ook de univariate resultaten van de categorische variant weer.

Vervolgens zijn de univariate analyses voor de overige factoren uitgevoerd en is de odds ratio berekend met bijbehorend 95% betrouwbaarheidsinterval. In bijlage B zijn de resultaten

<sup>9</sup> Bij deze vier variabelen vermoeden we dat het verband met Werkhervatting mogelijk niet-lineair is. De analyse met de categorische variant kan inzicht geven in dit niet-lineaire verband. We kunnen ook een niet lineaire functie schatten (bijv. een kwadratisch functie), dit heeft echter niet de voorkeur hier omdat de interpretatie lastiger wordt. Voor de categorische variant gebruiken we in eerste instantie de indeling die gebruikelijk is voor UWV. Echter, op basis van de multivariate analyses creëren we op basis van statistische gronden een categorische variant voor de variabelen Maximale duur recht op WW en Gemiddeld aantal arbeidsuren voor werkloosheid (zie sectie 3.5).

weergegeven waarbij de factoren zijn geordend op basis van belangrijkheid in termen van de hoogte van hun univariate bijdrage aan de voorspelling (van hoog naar laag op basis van de Nagelkerke  $R^2$ ). De modellen voor status werkhervatting op 12, negen en zes maanden laten dezelfde top tien zien van belangrijkste factoren (bijlage B, tabellen B.1, B.2, B.3). Dit zijn: Leeftijd, Maximale duur WW-recht, zowel continue als categorische variant, Visie op terugkeer naar werk, Aantal jaren in laatste functie, Aantal uren in staat te werken, Gezondheidsbeleving<sup>10</sup>, Algemeen werkvermogen, Acceptatiebereidheid m.b.t. een voltijd baan en Werkzoekgedrag m.b.t. direct contact werkgevers. Daarnaast laten de univariate analyses voor alle factoren zien dat de aard van het verband in de verwachte richting is op basis van de achterliggende theorie.

Tabel B.1 van bijlage B laat zien dat Leeftijd, Maximale duur WW-recht en Aantal jaren in laatste functie, een negatieve relatie hebben met werkhervatting binnen 12 maanden. We noemen deze factoren: de kansverkleinende factoren. Eén jaar toename in leeftijd zorgt voor een 0,94 keer zo grote kans op werkhervatting<sup>11</sup> binnen 12 maanden, m.a.w. de kans wordt kleiner met het ouder worden. Respondenten die langer dan 12 maanden recht op WW hebben, hebben een 0,27 keer zo grote kans op werkhervatting binnen 12 maanden dan respondenten die dit niet hebben. Respondenten die meer dan tien jaar in hun laatste functie hebben gewerkt hebben een 0,21 keer zo grote kans op werkhervatting binnen 12 maanden dan respondenten die een half jaar of minder in hun laatste functie hebben gewerkt.

De factoren Visie op terugkeer naar werk, Aantal uren in staat om te werken, Gezondheidsbeleving, Algemeen werkvermogen, Acceptatiebereidheid m.b.t. een voltijd baan en Werkzoekgedrag m.b.t. direct contact werkgevers, hebben een positieve relatie met werkhervatting binnen 12 maanden. We noemen deze factoren: de "kansvergroterende factoren". De schaal Visie op terugkeer naar werk loopt van 1, zeer mee oneens, tot 5, zeer mee eens, en is het gemiddelde van de scores op drie stellingen over de eigen verwachting om snel werk te vinden, bijv. "Ik verwacht snel zelf een baan te vinden" (zie bijlage A). Een hogere score op deze schaal geeft aan dat de werkzoekende een hogere verwachting heeft over het zelf vinden van werk. Eén punt hogere score op deze schaal verhoogt de kans op werkhervatting binnen 12 maanden met een vermenigvuldigingsfactor van 2,07. Dit geeft dus een verdubbeling van de kans.

Eén uur meer in staat zijn om te werken verhoogt de kans op werkhervatting met een vermenigvuldigingsfactor van 1,05. Eén punt hogere score op de schaal Gezondheidsbeleving verhoogt de kans op werkhervatting met een vermenigvuldigingsfactor van 2,13. De schaal Gezondheidsbeleving loopt van 1, zeer slecht, tot 5, zeer goed, en is het gemiddelde van vier vragen over gezondheid (zie bijlage A). Eén punt hogere score op de variabele Algemeen werkvermogen verhoogt de kans op werkhervatting met een vermenigvuldigingsfactor van 1,37. Deze variabele is de score op vraag 9a (zie bijlage A) en loopt van 1 (niet in staat om te werken) tot 10 (werkvermogen in beste periode). Eén punt hogere score op de variabele Acceptatiebereidheid m.b.t. een voltijd baan verhoogt de kans op werkhervatting met een vermenigvuldigingsfactor van 1,34. De score op deze acceptatiebereidheid (vraag 15c) ligt tussen 1 (zeker niet) en 5 (zeker wel). Eén punt hogere score op de variabele Werkzoekgedrag m.b.t. direct contact werkgevers verhoogt de kans

---

<sup>10</sup> In Werkverkenner 1.0 is deze factor gemeten met 3 aparte items, zie bijlage A.

<sup>11</sup> Strikt genomen gaat het hier om de "odds" op werkhervatting in plaats van de "kans" op werkhervatting. De "odds" is  $p / (1-p)$ . De kans is hieruit af te leiden, namelijk  $p = \text{odds} / (1+\text{odds})$ . Een toename van de odds betekent tevens een toename van de kans. Voor een uitgebreide toelichting, zie bijlage C.



op werkhervatting met een vermenigvuldigingsfactor van 2,00. De waarden op deze variabele liggen tussen 1 (bijna nooit) tot 4 (zeer vaak) en is het gemiddelde van de scores op verschillende vragen naar contact met werkgevers: de vragen 7b (sollicitatie gesprekken gehad), 7c (gebeld met) en 7d (werkgevers bezocht; zie bijlage A).

### 3.4.1 Voorspellende eigenschappen van het model met alleen de factor Leeftijd

**Conclusie:** Het logistisch regressiemodel met alleen de factor Leeftijd als predictor voorspelt de status werkhervatting binnen 12 maanden voor 64% van de respondenten correct. De globale fitmaten van het model zijn niet goed genoeg (AUC = 0,70; Brierscore > 0,20).

Uit de univariate analyses komt naar voren dat de factor Leeftijd de belangrijkste voorspeller is van status werkhervatting. In paragraaf 3.5 bekijken we de voorspellende eigenschappen van multivariate modellen (i.e. modellen met meerdere factoren). Om beter inzicht te krijgen hoe goed een model met alleen de factor Leeftijd voorspelt, gaan we nog wat nader in op de eigenschappen van dit model.

Hoe goed in het algemeen een logistisch regressiemodel voorspelt, bepalen we aan de hand van twee globale maten, de Area-Under-the-Curve (AUC) en de Brierscore. De AUC is een maat voor het gemiddelde onderscheidend vermogen van het model tussen de twee groepen, in ons geval tussen de groep werkhervatters binnen 12 maanden en de groep niet-werkhervatters binnen 12 maanden. Een model met een AUC tussen de 0,70 en 0,80 heeft een redelijk onderscheidend vermogen, onder de 0,70 is matig en boven de 0,80 is goed. De Brierscore geeft de gemiddelde voorspelfout weer. Het is een maat voor de gemiddelde discrepantie tussen de voorspelde kans en de werkelijk geobserveerde status. Bij een gelijke verdeling van de geobserveerde uitkomst, dus 50% heeft de status werkhervatters en 50% niet-werkhervatters, varieert de Brierscore tussen 0, perfecte voorspelling, en 0,25. Een Brierscore lager dan 0,20 wordt als goed beschouwd (Van Houwelingen & Putter, 2011).

Op basis van een afkapwaarde op de voorspelde kans op werkhervatting kunnen we de sensitiviteit, specificiteit en het percentage correct voorspeld bepalen. Bijvoorbeeld, als we een afkapwaarde nemen van 0,45, dan voorspellen we dat iedereen met een voorspelde kans hoger dan 0,45 behoort tot de categorie “werkhervatter binnen 12 maanden” en iedereen met een kans lager of gelijk aan 0,45 tot de categorie “niet-werkhervatter binnen 12 maanden”.

De *sensitiviteit* is het percentage werkhervatters binnen 12 maanden dat correct wordt voorspeld door het model. Oftewel, 1 minus de sensitiviteit is het percentage fout-negatieven (i.e. fout van de tweede soort, zie bijlage C).

De *specificiteit* is het percentage niet-werkhervatters binnen 12 maanden dat correct wordt voorspeld door het model. Oftewel, 1 minus de specificiteit is het percentage fout-positieven (i.e. fout van de eerste soort, zie bijlage C).

In onze analyses hebben we ofwel de sensitiviteit (SN) ofwel de specificiteit (SP) vastgezet op bepaalde waarden: 0,95, 0,90, 0,85, 0,80, 0,75 en 0,70 (zie bijlage C voor een uitgebreide uitleg). Bij deze vastgezette waarden hoort steeds een specifieke afkapwaarde op de voorspelde kans (zie tabel 3.3). Voor het bepalen van de voorspelkracht van het model, kijken we naar de afkapwaarde die resulteert in het hoogste percentage correct voorspeld en vervolgens kijken we naar de afkapwaarde die de hoogste som van sensitiviteit en specificiteit (SN + SP) bereikt. Door de hoogste percentages correct voorspeld van de



verschillende voorspelmodellen en de naïeve methode te vergelijken, bepalen we het beste voorspelmodel. Voor de toepassing van de Werkverkenner in de praktijk kan UWV andere afkapwaarden kiezen om klanten te selecteren voor intensievere dienstverlening. Deze keuze is afhankelijk van beleid van UWV, het budget dat beschikbaar is voor dienstverlening en arbeidsmarktontwikkelingen. Voor selectie van klanten kan UWV besluiten om een afkapwaarde te hanteren met de hoogste som van sensitiviteit en specificiteit (SN + SP). Dit betekent dat UWV dan streeft naar een balans tussen zo weinig mogelijk fout positieven en fout negatieven (zie bijlage C)<sup>12</sup>. Het kan voor de UWV praktijk ook wenselijk zijn om het percentage fout negatieven zo laag mogelijk te houden en dus de specificiteit zo hoog mogelijk. Hiermee kunnen we ervoor zorgen dat zo veel mogelijk niet-werkhervatters correct voorspeld worden. Daarom laten we ook de som van SN + 2SP zien (voor verdere toelichting, zie bijlage C).

De globale maten voor het model met alleen de factor Leeftijd geven aan dat het model op de grens ligt tussen een redelijk en matig onderscheidend vermogen (AUC = 0,70; tabel 3.3) en de gemiddelde voorspelfout is hoog (Brierscore > 0,20). Voor de voorspelling van status werkhervatting binnen 12 maanden is SN + SP het hoogst (128) wanneer we een afkapwaarde van 0,45 gebruiken op de voorspelde kans op werkhervatting. Van de werkzoekenden in de testset wordt dan 64% correct geclassificeerd.

Tabel 3.3: Voorspellende eigenschappen van het logistisch regressiemodel met alleen de factor Leeftijd voor de voorspelling van **status werkhervatting binnen 12 maanden** in de testset ( $n = 13.269$ ).

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct Voorspeld	SN + SP	SN + 2SP	AUC	Brier
20,6	94,8	0,74	56,4	115,3	210,1	0,700	0,220
30,4	90,2	0,70	59,2	120,6	210,7		
39,5	84,2	0,66	61,1	123,7	207,9		
44,7	80,1	0,63	61,8	124,9	205,0		
50,2	75,8	0,60	62,6	126,0	201,9		
57,4	69,2	0,55	63,1	126,6	195,8		
71,1	56,4	0,47	64,0	127,5	183,9		
<b>74,1</b>	<b>53,3</b>	<b>0,45</b>	<b>64,1</b>	<b>127,4</b>	<b>180,8</b>		
78,9	47,5	0,42	63,8	126,4	173,9		
85,6	38,0	0,37	62,6	123,5	161,5		
89,9	31,7	0,34	61,8	121,6	153,2		
94,5	22,7	0,30	59,8	117,2	139,8		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met de hoogste waarde voor %correct voorspeld heeft een donkerblauwe balk.

<sup>12</sup> We beperken ons hier tot het bekijken van 1 afkapwaarde. Voor het eindmodel (na stap 3) zullen we meerdere afkapwaardes bekijken.

### 3.5 Voorspelling werkhervatting met meerdere factoren samen

**Conclusie:** Het meervoudig logistisch regressiemodel met alle factoren samen voorspelt de status werkhervatting binnen 12 maanden voor 71% van de respondenten correct. Dit percentage is 69% binnen 9 maanden en 74% binnen 6 maanden. De toename in % correct voorspeld t.o.v. de naïeve methode is 19 procentpunt op 12 maanden, 12 procentpunt op 9 maanden en 4 procentpunt op 6 maanden. We richten ons voor de interpretatie op de modellen van 12 en 9 maanden. Het model van 6 maanden laten we buiten beschouwing omdat het niet goed genoeg is (i.e. slechts 36% van de werkelijke werkhervatters wordt goed voorspeld). Zou UWV kiezen voor een afkapwaarde met de hoogste som van sensitiviteit en specificiteit (SN + SP), dan voorspelt het model op 6 maanden met alle factoren samen slechter dan de naïeve methode.

We onderzochten tevens welke factoren in het meervoudig logistisch regressiemodel een verband in de op de theorie gebaseerde / theoretisch verwachte richting lieten zien en of we het aantal factoren konden reduceren. Het meest spaarzame, interpreteerbare model voor de voorspelling van status werkhervatting binnen 12 maanden bevat 18 factoren en de voorspelkracht is vrijwel gelijk aan het model met alle factoren (70% correct voorspeld; AUC = 0,78; Brier = 0,191). Globaal kunnen we spreken van een redelijk discriminerend ( $0,70 < \text{AUC} < 0,80$ ) en goed passend model (Brier  $< 0,20$ ). Zeven van de 18 factoren in dit nieuwe model zitten ook in Werkverkenner 1.0 (merk hierbij op dat 3 factoren van Werkverkenner 1.0 zijn samengenomen tot 1 factor “gezondheidsbeleving” in het nieuwe model, dus vanuit Werkverkenner 1.0 bekeken, kunnen we ook stellen dat negen factoren overeenkomen). De belangrijkste factoren in het spaarzame model zijn: Leeftijd, Aantal jaren in laatste functie, Visie op terugkeer naar werk, Bemiddelingsberoep (soort), Samenstelling huishouden, Maximale duur WW-recht en Bedrijfstak.

Het meest spaarzame, interpreteerbare model voor de voorspelling van status werkhervatting binnen 9 maanden bevat 17 factoren en de voorspelkracht is vrijwel gelijk aan het model met alle factoren (69% correct voorspeld; AUC = 0,76; Brier = 0,191). Globaal kunnen we ook spreken van een redelijk discriminerend ( $0,70 < \text{AUC} < 0,80$ ) en goed passend model (Brier  $< 0,20$ ). Vijftien van deze 17 factoren zijn gelijk aan het model van 12 maanden.

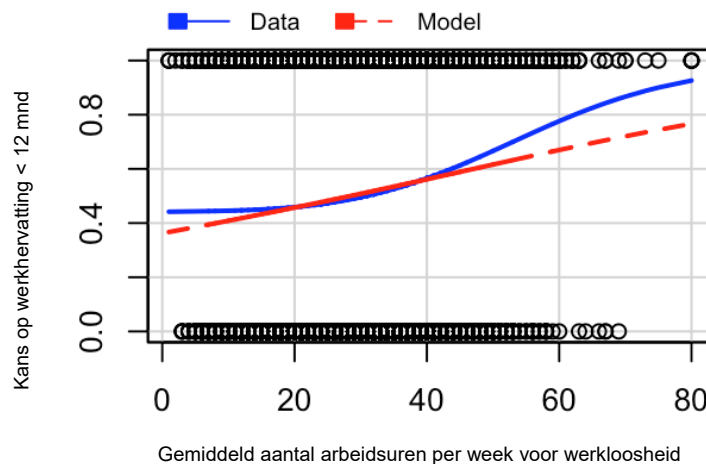
Om deelvragen 1b en 1c te beantwoorden voeren we meervoudige logistische regressie analyses uit. In een dergelijke analyse wordt de invloed van meerdere factoren samen op de uitkomstmaat geschat. Hierdoor kunnen we over het algemeen nauwkeuriger en beter voorspellen. Het resultaat van de meervoudige logistische regressie analyse is een gewogen som van de factoren in het model; deze gewogen som geeft de voorspelde  $\log(\text{odds})$  op werkhervatting<sup>13</sup>. De geschatte gewichten bij de factoren passen zo goed mogelijk bij de gegevens die zijn gebruikt voor het model. Om nu te bepalen hoe goed het model erin slaagt om voor nieuwe werkzoekenden de status werkhervatting te voorspellen, maken we bij de analyses gebruik van de ontwikkelset (om het model te schatten), de validatieset (om het model te vereenvoudigen) en de testset (om het model te testen en dus de voorspelkracht voor nieuwe werkzoekenden te bepalen). Bij de analyses hebben we verschillende stappen doorlopen (voor een uitgebreide beschrijving van de stappen en de gebruikte statistieken, zie bijlage C). Als initiële stap (stap 0), hebben we eerst de belangrijkste aannames van het meervoudige model gecheckt: multicollineariteit ( $r > 0,80$  en VIF waardes  $< 5$ ) en lineariteit (voor continue predictoren: *marginal model plots*). Bij twee paren van factoren was er sprake van multicollineariteit: Gemiddeld aantal arbeidsuren voor werkloosheid met Werkloosheidsomvang in uren ( $r = 0,87$ ) en Leeftijd met Maximale duur WW-recht ( $r = 0,91$ ). Als we een model schatten met deze vier factoren erin (en de overige factoren) dan zijn de

<sup>13</sup> De  $\log(\text{odds})$  is om te zetten in de kans op werkhervatting ( $p$ ). Stel de  $\log(\text{odds}) = x$ . Dan is  $p = \frac{\exp^x}{1 + \exp^x}$ .

VIF waarden van deze factoren hoger dan 5. Daarom hebben we per paar 1 factor geselecteerd. Wat betreft het eerste paar hebben we alleen Gemiddeld aantal arbeidsuren voor werkloosheid meegenomen in de analyse. Wat betreft het tweede paar hebben we Maximale duur WW-recht gedichotomiseerd op de volgende manier. Degenen met een maximale duur WW-recht langer dan 12 maanden (of 9 maanden of 6 maanden, afhankelijk van de uitkomstmaat) krijgen een waarde 1 en degenen met een maximale duur WW-recht van 12 maanden of korter krijgen een waarde 0. De gedichotomiseerde variabele had een veel lagere correlatie met Leeftijd ( $r = 0,65$ ). De VIF waarden van het model zonder de factor Werkloosheidsomvang en met de gedichotomiseerde factor maximale duur WW-recht zijn allen onder de 5.

Eén marginal model plot (zie figuur 3.1) gaf aan dat het verband niet-lineair was voor Gemiddeld aantal arbeidsuren voor werkloosheid met status werkhervatting (binnen 12, 9 en 6 maanden; dus voor elk van de meetmomenten). De discrepantie tussen de rode en blauwe lijn in de figuur duidt op een niet-lineaire relatie. Tot ca. 24 uur is de blauwe lijn vlak en vanaf 24 uur sterk stijgend. We hebben vervolgens deze variabele gecategoriseerd op basis van de figuur in: 24 uur of minder, tussen 25 en 32 uur, meer dan 32 uur.

In de rest van deze paragraaf beschrijven we de resultaten van de volgende drie stappen: stap 1 het meervoudige logistische regressiemodel met alle factoren samen (onderzoeksvraag 1b, zie par. 3.1); stap 2 het interpreteerbare meervoudige model (d.w.z.. zonder de factoren die moeilijk te interpreteren zijn; dit zullen we verderop toelichten); stap 3 het interpreteerbare, meest spaarzame meervoudige model (onderzoeksvraag 1c).



Figuur 3.1: Marginale model plot van de factor Gemiddeld aantal arbeidsuren voor werkloosheid met de kans op werkhervatting binnen 12 maanden

### 3.5.1 Resultaten stap 1: hoe voorspellend zijn alle factoren samen

Tabellen 3.4a t/m 3.4c geven de voorspellende eigenschappen weer van het logistisch regressiemodel met alle factoren samen voor de voorspelling van werkhervatting binnen 12, 9 en 6 maanden, op dezelfde manier als is toegelicht in par. 3.4.1.

Tabel 3.4a: Voorspellende eigenschappen van het meervoudig logistisch regressiemodel met alle factoren samen voor de voorspelling van **status werkhervatting binnen 12 maanden**;  $n = 13.269$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct Voorspeld	SN + SP	SN + 2SP	AUC	Brier
27,7	95,0	0,80	60,2	122,7	217,7	0,783	0,189
40,8	90,0	0,72	64,5	130,8	220,8		
50,8	85,0	0,66	67,3	135,8	220,8		
58,5	80,0	0,61	68,9	138,5	218,5		
65,1	75,0	0,56	69,9	140,1	215,1		
70,5	70,0	0,52	70,3	140,6	210,6		
70,0	70,6	0,52	70,3	140,6	211,1		
<b>75,0</b>	<b>66,4</b>	<b>0,49</b>	<b>70,9</b>	<b>141,4</b>	<b>207,8</b>		
80,0	61,0	0,45	70,9	141,0	202,1		
85,0	54,3	0,40	70,2	139,3	193,6		
90,0	46,2	0,34	68,8	136,2	182,3		
95,0	34,6	0,26	65,9	129,6	164,3		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met de hoogste waarde voor % correct voorspeld (en vervolgens voor SN + SP) heeft een donkerblauwe balk.

Tabel 3.4b: Voorspellende eigenschappen van het meervoudig logistisch regressiemodel met alle factoren samen voor de voorspelling van **status werkhervatting binnen 9 maanden**;  $n = 13.269$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
26,3	95,0	0,72	65,4	121,3	216,3	0,772	0,192
38,4	90,0	0,64	67,8	128,4	218,4		
49,0	85,0	0,58	69,5	134,0	219,0		
56,7	80,0	0,52	70,0	136,7	216,8		
63,3	75,0	0,48	70,0	138,3	213,3		
68,9	70,0	0,44	69,5	138,9	208,9		
<b>70,0</b>	<b>69,1</b>	<b>0,43</b>	<b>69,5</b>	<b>139,1</b>	<b>208,1</b>		
75,0	64,2	0,40	68,8	139,2	203,3		
80,0	59,0	0,36	68,0	139,0	198,0		
85,0	52,5	0,32	66,5	137,5	189,9		
90,0	44,1	0,27	63,8	134,1	178,1		
95,0	32,7	0,21	59,5	127,7	160,4		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met de hoogste waarde voor % correct voorspeld (en vervolgens hoogste SN + SP) heeft een donkerblauwe balk.

Tabel 3.4c: Voorspellende eigenschappen van het meervoudig logistisch regressiemodel met alle factoren samen voor de voorspelling van **status werkhervatting binnen 6 maanden**;  $n = 13.269$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
24,3	95,0	0,58	73,7	119,3	214,3	0,758	0,174
<b>36,2</b>	<b>90,0</b>	<b>0,49</b>	<b>73,8</b>	<b>126,2</b>	<b>216,2</b>		
46,3	85,0	0,43	73,3	131,3	216,3		
54,4	80,0	0,38	72,3	134,4	214,4		
60,7	75,0	0,34	70,7	135,7	210,7		
66,8	70,0	0,31	69,0	136,8	206,8		
70,0	66,9	0,29	67,8	136,9	203,8		
75,0	61,5	0,26	65,6	136,5	198,0		
80,0	55,9	0,24	63,2	135,9	191,8		
85,0	49,3	0,21	60,1	134,3	183,6		
90,0	41,5	0,17	56,1	131,5	173,0		
95,0	30,8	0,13	50,2	125,8	156,7		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met de hoogste waarde voor % correct voorspeld heeft een donkerblauwe balk.

Tabellen 3.4a t/m 3.4c laten zien dat het onderscheiden vermogen (de AUC) van de modellen voor alle drie de tijdstippen redelijk is. De AUC neemt af van 0,78 op 12 maanden, naar 0,77 op 9 maanden en 0,76 op 6 maanden. De gemiddelde voorspelfout voor alle drie de modellen is goed. De Brierscore is vrijwel gelijk op 12 en 9 maanden (0,19) en lager op 6 maanden (0,17). Voor het vergelijken van de voorspelkracht van de modellen, kijken we naar de afkapwaarde die het hoogste percentage correct voorspeld bereikt (en als twee afkapwaardes tot ditzelfde percentage leiden, kijken we vervolgens naar SN + SP; zie bijlage C)<sup>14</sup>.

Voor de voorspelling van status werkhervatting *binnen 12 maanden* is het percentage correct voorspeld het hoogst wanneer we een afkapwaarde van 0,49 gebruiken op de voorspelde kans op werkhervatting (zie donkerblauwe balk in tabel 3.4a). Dit betekent dat werkzoekenden met een voorspelde kans van 0,49 of hoger de voorspelde status werkhervatter binnen 12 maanden krijgen en lager dan 0,49 krijgen de voorspelde status niet-werkhervatter binnen 12 maanden. We vergelijken deze voorspelde status met de werkelijk geobserveerde status van de werkzoekenden in de testset en 70,9% van hen wordt dan correct geclassificeerd. Vergelijken we dit met de naïeve methode die 52% correct voorspelde, dan kunnen we afgerond 19%-punt meer correct voorspellen met dit model.

Voor de voorspelde kans op werkhervatting *binnen 9 maanden* is het percentage correct voorspeld het hoogst wanneer we een afkapwaarde van 0,43 gebruiken (tabel 3.4b). Van de werkzoekenden in de test set wordt 69,5% correct geclassificeerd. Vergelijken we dit met de naïeve methode die 57% correct voorspelde, dan kunnen we afgerond 13%-punt beter voorspellen met dit model.

Voor de voorspelde kans op werkhervatting *binnen 6 maanden* is het percentage correct voorspeld het hoogst wanneer we een afkapwaarde van 0,49 gebruiken (tabel 3.4c). Van de

<sup>14</sup> We beperken ons hier tot het bekijken van 1 afkapwaarde. Voor het eindmodel (na stap 3) zullen we meerdere afkapwaardes bekijken.

werkzoekenden in de test set wordt 73,8% correct geclassificeerd. Hierbij merken we op dat de specificiteit weliswaar hoog is bij deze afkapwaarde (90%), maar de sensitiviteit is erg laag (36,2%). Vergelijken we dit model met de naïeve methode die 70% correct voorspelde, dan kunnen we afgerond 4%-punt beter voorspellen.

We richten ons voor de interpretatie van de resultaten op de modellen van 12 en 9 maanden. Het model van 6 maanden laten we buiten beschouwing om de volgende redenen: 1) de toegevoegde waarde t.o.v. de naïeve methode is klein, slechts 4% en 2) de sensitiviteit van het model is laag, slechts 36% van de werkelijke werkhervatters binnen 6 maanden wordt correct geclassificeerd. Voor de praktijk van UWV zou dit model de volgende implicaties hebben. Als iedereen met een voorspelde kans lager dan 0,49 (met dit model) wordt uitgenodigd, dan betekent dit dat veel werkzoekenden worden uitgenodigd voor extra dienstverlening die zonder dienstverlening ook wel aan het werk komen binnen 6 maanden (nl. 64%); bovendien zou in totaal 74% van alle werkzoekenden worden uitgenodigd<sup>15</sup>. Voor selectie van klanten is dit voorspelmodel niet goed genoeg. Zou UWV kiezen voor een afkapwaarde met de hoogste som van sensitiviteit en specificiteit (SN + SP), dan voorspelt het model 67,8% correct (zie tabel 3.4c) en daarmee slechter dan de naïeve methode.

Op zowel 12 als 9 maanden laten een zevental factoren in de meervoudige modellen (bijlage D, tabellen D.1 en D.2) een andere richting van de regressiecoëfficiënt zien dan in het corresponderende univariate model (bijlage B, tabellen B1 en B2). De richting is anders dan we op basis van de theorie verwachten (zie hoofdstuk 1). Het gaat om de factoren Werkzoekattitude m.b.t. voordeel en aangenaamheid van werk zoeken, Werkzoekattitude m.b.t. nut en noodzaak van werk zoeken, Externe variabele attributie, Acceptatiebereidheid m.b.t. onwenselijk werk, Self-efficacy m.b.t. sollicitatievoorbereiding, Gemiddeld aantal arbeidsuren voor werkloosheid en Subjectieve norm familie en partner (NB. Voor het model van 6 maanden zijn dit dezelfde factoren m.u.v. Subjectieve norm). De effect groottes van deze factoren zijn 0,003 of lager (zie de kolom met de relatieve toename in Nagelkerke  $R^2$  in tabellen D.1 en D.2); dit betekent dat de unieke bijdrage van de factoren aan de voorspelling relatief klein is. Omdat we de tegengestelde richtingen van de zeven factoren niet kunnen verklaren met een achterliggend theoretisch model over bijvoorbeeld mediatie-effecten en omdat de bijdrage van de factoren aan de voorspelling erg klein is, hebben we vervolgens in stap 2 deze niet goed interpreteerbare factoren buiten beschouwing gelaten in de analyses.

### 3.5.2 Resultaten stap 2: Het beter interpreteerbare model

We bekijken de voorspellende eigenschappen van de modellen zonder de niet-interpreteerbare factoren om te zien of we een min of meer gelijke voorspelling krijgen als de modellen met alle factoren erin (stap 1). De tabellen in bijlage E geven de voorspellende eigenschappen van de modellen weer zonder de niet-interpreteerbare factoren. Deze voorspellende eigenschappen verschillen nauwelijks van die van het vorige model van stap 1. Dit betekent dat we verder kunnen gaan naar stap 3 met alleen de goed interpreteerbare verbanden in het model.

### 3.5.3 Resultaten stap 3: Het meest spaarzame, interpreteerbare model

We nemen het model van stap 2 als uitgangspunt en passen vervolgens de verschillende stapsgewijze selectieprocedures toe (zie bijlage C voor toelichting). De selectieprocedure die

---

<sup>15</sup> Het % werkhervatters binnen zes maanden is gelijk aan 30% (zie tabel 3.1), dus het % niet-werkhervatters is 70%. De sensitiviteit van de voorspelregel met afkapwaarde 0,49 is 36% en de specificiteit is 90%. Dit impliceert dat met deze regel in totaal 74% (i.e.  $0,36 * 0,30 + 0,90 * 0,70$ ) van alle werkzoekenden zou worden uitgenodigd, dat is dus 7 op de 10 werkzoekenden.

resulteert in het model met het minste aantal factoren en met goede voorspellende eigenschappen in de validatieset heeft de voorkeur. Voor de voorspelling van zowel status werkhervatting binnen 12 maanden als binnen 9 maanden is dit het model dat wordt geselecteerd met de BIC methode. De drie varianten van de BIC methode (forward, backward en stapsgewijze selectie) resulteerden in hetzelfde model. Het uiteindelijke model voor status werkhervatting binnen 12 maanden bevat 18 van de in totaal 37 factoren en het model voor status werkhervatting binnen 9 maanden bevat 17 factoren.

Tabel 3.5a: Voorspellende eigenschappen van het meest spaarzame, interpreteerbare model voor de voorspelling van **status werkhervatting binnen 12 maanden**;  $n = 13.269$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
25,7	95,0	0,80	59,1	120,7	215,7	0,778	0,191
40,1	90,0	0,72	64,2	130,1	220,1		
49,2	85,0	0,66	66,5	134,2	219,2		
57,3	80,0	0,61	68,3	137,3	217,3		
63,7	75,0	0,57	69,1	138,7	213,7		
69,1	70,0	0,53	69,5	139,1	209,1		
70,0	69,2	0,52	69,6	139,2	208,3		
75,0	64,7	0,49	70,0	139,7	204,4		
<b>80,0</b>	<b>59,9</b>	<b>0,45</b>	<b>70,3</b>	<b>139,9</b>	<b>199,9</b>		
85,0	53,7	0,40	69,9	138,7	192,5		
90,0	46,1	0,34	68,8	136,1	182,1		
95,0	34,3	0,27	65,7	129,3	163,6		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met de hoogste waarde voor % correct voorspeld heeft een donkerblauwe balk.

Tabel 3.5b: Voorspellende eigenschappen van het meest spaarzame, interpreteerbare model voor de voorspelling van **status werkhervatting binnen 9 maanden**;  $n = 13.269$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
25,6	95,0	0,72	65,1	120,6	215,6	0,765	0,194
37,8	90,0	0,64	67,6	127,8	217,8		
47,7	85,0	0,58	69,0	132,7	217,7		
<b>55,0</b>	<b>80,0</b>	<b>0,53</b>	<b>69,3</b>	<b>135,0</b>	<b>215,0</b>		
61,5	75,0	0,49	69,2	136,4	211,4		
67,6	70,0	0,45	69,0	137,6	207,6		
70,0	67,6	0,43	68,6	137,6	205,1		
75,0	62,6	0,40	67,9	137,6	200,1		
80,0	57,1	0,36	66,9	137,1	194,2		
85,0	51,1	0,32	65,7	136,1	187,1		
90,0	43,7	0,28	63,6	133,7	177,4		
95,0	31,5	0,21	58,8	126,5	157,9		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met de hoogste waarde voor % correct voorspeld heeft een donkerblauwe balk.

Vervolgens hebben we de voorspellende eigenschappen van deze spaarzame modellen bepaald in de testset (tabellen 3.5a en 3.5b). Het model heeft dan een AUC waarde van

0,778 voor 12 maanden en 0,765 voor 9 maanden. Voor 12 maanden geeft een afkapwaarde van 0,45 het hoogste percentage correct voorspeld (tabel 3.5a). Voor 9 maanden geeft een afkapwaarde van 0,53 het hoogste percentage correct voorspeld (tabel 3.5b).

Tabel 3.5c: Kruistabel met de voorspelde en geobserveerde waarden van de status werkhervatting binnen 12 maanden met een gebruikte afkapwaarde van 0,45 in de test set ( $n = 13.269$ ).

		Geobserveerde status werkhervatting		Rij totaal
		0 (niet-werkhervatter)	1 (werkhervatter)	
Voorspelde status werkhervatting ( $\hat{p} > 0,45$ )	0 (niet-werkhervatter)	3.860 (29%)	1.390 (11%)	5.250 (40%)
	1 (werkhervatter)	2.541 (19%)	5.478 (41%)	8.019 (60%)
	Kolom totaal	6.401 (48%)	6.868 (52%)	13.269 (100%)

*Noot.* De percentages in de cellen geven het percentage van de totale groep ( $n = 13.269$ ) weer. De sensitiviteit is gelijk aan  $5.478 / 6.868 = 80\%$  (zie ook de donkerblauwe regel van tabel 3.5a). De specificiteit is gelijk aan  $3.860 / 6.401 = 60\%$ .

Tabel 3.5d: Kruistabel met de voorspelde en geobserveerde waarden van de status werkhervatting binnen 12 maanden met een gebruikte afkapwaarde van 0,72 in de test set ( $n = 13.269$ ).

		Geobserveerde status werkhervatting		Rij totaal
		0 (niet-werkhervatter)	1 (werkhervatter)	
Voorspelde status werkhervatting ( $\hat{p} > 0,72$ )	0 (niet-werkhervatter)	5.757 (43%)	4.110 (31%)	9.867 (74%)
	1 (werkhervatter)	644 (5%)	2.758 (21%)	3.402 (26%)
	Kolom totaal	6.401 (48%)	6.868 (52%)	13.269 (100%)

*Noot.* De percentages in de cellen geven het percentage van de totale groep ( $n = 13.269$ ) weer. De sensitiviteit is gelijk aan  $2.758 / 6.868 = 40\%$  (zie ook de regel met afkapwaarde 0,72 in tabel 3.5a). De specificiteit is gelijk aan  $5.757 / 6.401 = 90\%$ .

Zoals eerder uitgelegd (zie 3.4.1) wordt de afkapwaarde gebruikt om de voorspelde status werkhervatting te bepalen. Dit gaat op de volgende manier: als de voorspelde kans op werkhervatting binnen 12 maanden hoger is dan de afkapwaarde (van bijv. 0,45), dan geven we de werkzoekende als voorspelde status werkhervatter en anders wordt de voorspelde status niet-werkhervatter (tabel 3.5c). Dit resulteert in een correcte voorspelling voor  $3.860 + 5.478 = 9.338$  van de personen in de testset ( $29\% + 41\% = 70\%$ ). In de praktijk kan UWV besluiten om voor de groep met een voorspelde status niet-werkhervatter extra dienstverlening in te zetten om ervoor te zorgen dat zij weer aan het werk gaan. Tabel 3.5c laat zien dat bij een afkapwaarde van 0,45 in totaal 5.250 (40%) werkzoekenden in aanmerking komen voor dienstverlening (zie ook bijlage C). Gebruiken we een hogere afkapwaarde (bijvoorbeeld 0,72, dit is de waarde bij de hoogste  $SN + 2SP$ ), dan komen er in totaal 9.867 (74%) werkzoekenden in aanmerking voor extra dienstverlening (zie tabel 3.5d).

De fout positieven zijn de werkzoekenden met een geobserveerde status niet-werkhervatter maar met een voorspelde status werkhervatter, dus, met andere woorden werkzoekenden die extra dienstverlening zouden kunnen gebruiken, maar deze niet krijgen. Dit aantal fout positieven is met een afkapwaarde van 0,72 veel lager, namelijk 644 (5% van het totaal; tabel 3.5d) dan met een afkapwaarde van 0,45, namelijk 2.541 (19% van het totaal; tabel 3.5c).



De fout negatieven zijn werkzoekenden met een geobserveerde status werkhervatter maar met een voorspelde status niet-werkhervatter, met andere woorden werkzoekenden die zonder extra dienstverlening ook wel aan het werk komen. Dit aantal fout negatieven is met een afkapwaarde van 0,72 hoger, namelijk 4.110 (31% van het totaal; tabel 3.5d) dan met een afkapwaarde van 0,45, namelijk 1.390 (11% van het totaal; tabel 3.5c). Fout negatieven en fout positieven kunnen worden gezien als communicerende vaten.

Tabel 3.6 geeft het spaarzame model weer voor status werkhervatting binnen 12 maanden. De verklarende eigenschap (Nagelkerke  $R^2$ ) van het totale model is 0,313. De kansvergrotenende factoren zijn in groen afgebeeld en de kansverkleinende in rood. Een categorische factor (bijv. Samenstelling Huishouden) kan zowel kansvergrotenend zijn als kansverkleinend. Dit hangt af van de categorie die als referentie wordt genomen. De keuze van de referentiecategorie is arbitrair en heeft verder geen invloed op de waarde van de belangrijkheid van de factor  $\Delta R^2$ ; in hoofdstuk 6 gaan we verder in op de verschillen tussen de categorieën). We kiezen de referentiecategorie zodanig dat het effect van alle andere categorieën dezelfde richting op gaat. Bijvoorbeeld, voor Samenstelling huishouden nemen we 'woont alleen' als referentie categorie. We zien dat alle andere categorieën t.o.v. woont alleen de kans op werkhervatting vergroten of geen significant verschil laten zien met de referentiecategorie (een niet significante categorie is in lichtblauw weergegeven in tabel 3.6). Daarom beschouwen we de factor Samenstelling huishouden als kansvergrotenend. Hieronder gaan we eerst in op de kansvergrotenende factoren (13 in totaal), daarna op de kansverkleinende factoren (5 in totaal).

De dertien *kansvergrotenende factoren* zijn (in volgorde van de hoogte van de relatieve  $R^2$ , d.w.z. de unieke bijdrage van de factor aan de voorspelling): Visie op terugkeer naar werk, Samenstelling huishouden, Bedrijfstak, Werkzoekgedrag m.b.t. direct contact met werkgevers, Gezondheidsbeleving, Balans voor- en nadelen niet-werken, Acceptatiebereidheid m.b.t. voltijd baan, Algemeen werkvermogen, Aantal uren in staat te werken, Opleidingsniveau, Acceptatiebereidheid m.b.t. variabele en/of afwijkende tijden en Inkomsten uit werk naast WW. We zullen iets verder ingaan op de interpretatie van deze factoren:

- 1) Visie op terugkeer naar werk: Eén punt hogere score op de schaal Visie op terugkeer naar werk verhoogt de kans op werkhervatting binnen 12 maanden met een vermenigvuldigingsfactor van 1,39.
- 2) Bemiddelingsberoep (soort)<sup>16</sup>: Werkzoekenden uit de agrarische beroepen hebben een 2,04 keer zo grote kans op werkhervatting binnen 12 maanden dan werkzoekenden uit de economisch-administratieve beroepen. Ook werkzoekenden uit de technische en industrieberoepen en uit de transportberoepen hebben een, respectievelijk, 1,77 en 1,41 keer zo grote kans op werkhervatting binnen 12 maanden als werkzoekenden uit de economisch-administratieve beroepen. Werkzoekenden uit de informatica, medisch/paramedische beroepen en dienstverlenende en verzorgende beroepen hebben 1,22, 1,28 en 1,26 keer zo grote kans op werkhervatting binnen 12 maanden als werkzoekenden uit de economisch-administratieve beroepen. Voor de overige bemiddelingsberoepen zijn de verschillen niet significant met de economisch-administratieve beroepen.
- 3) Samenstelling huishouden: Werkzoekenden die getrouwd zijn of samen wonen met thuiswonende kinderen van 7 jaar en ouder hebben een 1,55 keer zo grote kans op

---

<sup>16</sup> De invloed van deze factor is conjunctuurgevoelig en dus vluchtig van aard. Het is van belang dat de gewichten van de categorieën van deze factor in de voorspelregel regelmatig worden geactualiseerd (zie ook de aanbevelingen in de samenvatting).

werkhervatting binnen 12 maanden als werkzoekenden die alleen wonen. Voor de groep werkzoekenden die getrouwd zijn of samen wonen zonder thuiswonende kinderen is deze kans een factor 1,23 keer groter. Eveneens hebben mannen die getrouwd zijn of samen wonen met het jongste thuiswonende kind tussen 0-6 jaar een 1,24 keer zo grote kans op werkhervatting binnen 12 maanden als werkzoekenden die alleen wonen. De overige categorieën verschillen niet significant van werkzoekenden die alleen wonen.

- 4) Bedrijfstak<sup>15</sup>: Werkzoekenden uit de bedrijfstak bouw en hout hebben een 1,85 keer zo grote kans op werkhervatting binnen 12 maanden als werkzoekenden uit de (financiële) dienstverlening. Ook werkzoekenden uit de bedrijfstak uitzendbedrijf en uit de bedrijfstak landbouw, visserij en voeding hebben respectievelijk een 1,62 en 1,38 keer zo grote kans op werkhervatting binnen 12 maanden als werkzoekenden uit de (financiële) dienstverlening. De overige bedrijfstakken hebben ook een grotere kans op werkhervatting vergeleken met werkzoekenden uit de (financiële) dienstverlening, maar de factor is kleiner (1,30).
- 5) Werkzoekgedrag m.b.t. direct contact werkgevers: Eén punt hogere score op de variabele Werkzoekgedrag m.b.t. direct contact werkgevers verhoogt de kans op werkhervatting met een vermenigvuldigingsfactor van 1,27.
- 6) Gezondheidsbeleving: Eén punt hogere score op de schaal Gezondheidsbeleving verhoogt de kans op werkhervatting binnen 12 maanden met een vermenigvuldigingsfactor van 1,29.
- 7) Balans voor- en nadelen niet-werken: Werkzoekenden die meer nadelen zien van niet-werken dan voordelen hebben een 1,23 keer zo grote kans op werkhervatting als degenen die evenveel voor- als nadelen zien. Werkzoekenden die minder nadelen zien van niet-werken dan voordelen verschillen niet significant van degenen die evenveel voor- als nadelen zien.
- 8) Acceptatiebereidheid m.b.t. voltijdbaas: Eén punt hogere score verhoogt de kans op werkhervatting met een vermenigvuldigingsfactor van 1,08. De score op deze variabele varieert van 1 (zeker niet) tot 5 (zeker wel), zie vraag 15c in bijlage A.
- 9) Algemeen werkvermogen: Eén punt hogere score op de vraag naar Algemeen werkvermogen (vraag 9a, zie bijlage A) verhoogt de kans op werkhervatting binnen 12 maanden met een vermenigvuldigingsfactor van 1,07. De score op deze variabele varieert van 1 (niet in staat om te werken) tot 10 (werkvermogen in uw beste periode).
- 10) Aantal uren in staat te werken: Eén uur meer in staat om te werken verhoogt de kans op werkhervatting met een vermenigvuldigingsfactor van 1,01.
- 11) Opleidingsniveau: Werkzoekenden met een opleidingsniveau in de midden categorie hebben een 1,19 keer zo grote kans op werkhervatting als werkzoekenden met een laag opleidingsniveau. Werkzoekenden met een hoog opleidingsniveau hebben een 1,10 keer zo grote kans op werkhervatting als werkzoekenden met een laag opleidingsniveau.
- 12) Acceptatiebereidheid m.b.t. variabele en/of afwijkende tijden: Eén punt hogere score verhoogt de kans op werkhervatting met een vermenigvuldigingsfactor van 1,06. De score op deze variabele varieert van 1 (zeker niet) tot 5 (zeker wel) en is een combinatie van vraag 15a en 15b (bijlage A).

- 13) Inkomsten uit werk naast WW: Werkzoekenden die inkomsten uit werk hebben naast de WW-uitkering hebben een 1,17 keer zo grote kans op werkherhvatting als werkzoekenden die geen inkomsten uit werk hebben naast de WW-uitkering<sup>17</sup>.

De vijf *kansverkleinende factoren* zijn (in volgorde van de hoogte van de relatieve  $R^2$ , d.w.z. de unieke bijdrage van de factor aan de voorspelling, zie tabel 3.6): Leeftijd, Aantal jaren in laatste functie, Maximale duur WW-recht, WIA 35-min-indicatie en Problemen met de Nederlandse taal. We zullen ook verder ingaan op de interpretatie van deze factoren:

- 1) Leeftijd: Eén jaar ouder “verhoogt” de kans op werkherhvatting binnen 12 maanden met een vermenigvuldigingsfactor van 0,96. De kans op werkherhvatting wordt dus kleiner naarmate men op oudere leeftijd instroomt in de WW.
- 2) Aantal jaren in laatste functie: Werkzoekenden die meer dan tien jaar in hun laatste functie hebben gewerkt, hebben een 0,46 keer zo grote kans op werkherhvatting als werkzoekenden die een half jaar of minder in hun laatste functie hebben gewerkt (de kans wordt dus kleiner).
- 3) Maximale duur WW-recht: Werkzoekenden die langer dan 12 maanden recht hebben op een WW-uitkering hebben een 0,59 keer zo grote kans op werkherhvatting als degenen die tussen de drie en 12 maanden recht hebben op WW.
- 4) WIA 35-min-indicatie: WW'ers met een WIA 35-min-indicatie hebben een 0,50 keer zo grote kans op werkherhvatting als WW'ers zonder WIA 35-min-indicatie.
- 5) Problemen met de Nederlandse taal: Werkzoekenden die aangeven dat ze een beetje problemen met de Nederlandse taal hebben, hebben een 0,76 keer zo grote kans op werkherhvatting als degenen die aangeven dat ze geen problemen hebben met de Nederlandse taal. Werkzoekenden die aangeven dat ze veel problemen met de Nederlandse taal hebben of helemaal geen Nederlands kunnen, hebben een 0,63 keer zo grote kans op werkherhvatting als degenen die aangeven dat ze geen problemen hebben met de Nederlandse taal.

Het spaarzame model voor 9 maanden is weergegeven in bijlage F. De Nagelkerke  $R^2$  van dit model is 0,281. Vijftien van de in totaal 17 factoren komen overeen met het model voor 12 maanden (tabel 3.6). In vergelijking met het spaarzame model voor 12 maanden is in dit model het Bemiddelingsberoep (niveau) wel voorspellend en Problemen met de Nederlandse taal niet.

---

<sup>17</sup> Plausibele verklaringen voor de richting van het effect van deze factor zijn als volgt: als je al gedeeltelijk betaald werk verricht, is het aannemelijk dat je sneller helemaal het werk hervat. Bijvoorbeeld door uitbreiding van uren bij de huidige werkgever. Een andere mogelijke verklaring is dat mensen die nog inkomsten uit werk hebben, ook geen 'gat in hun CV krijgen'. Voor werkgevers waar zij solliciteren komen zij dus over als baanveranderaars en niet als werklozen.)

Tabel 3.6: Resultaten voor de voorspelling van status werkhervatting binnen 12 maanden met het meest spaarzame, interpreteerbare meervoudige logistisch regressiemodel (op de ontwikkelset:  $N = 26.541$ ). De factoren zijn geordend in volgorde van belangrijkheid op basis van de verklaarde variantie (relatieve toename in Nagelkerke  $R^2$ ). Kansverkleinende factoren zijn in lichtrood weergegeven. Kansvergrotenende factoren zijn in groen weergegeven. De Nagelkerke  $R^2$  van het totale model = 0,313.

WV	Nr	Factor	OR	BI		b	SE	z	p	$\Delta R^2$
				2,5%	97,5%					
1.0										
X	26	Leeftijd	0,96	0,95	0,96	-0,05	0,00	-23,09	< 0,001	0,031
X	3	Aantal jaren in laatste functie: 0 - 0,5 jaar	1,00							0,017
		Aantal jaren in laatste functie: meer dan 0,5 - 1 jaar	0,89	0,79	1,00	-0,12	0,06	-2,03	0,042	
		Aantal jaren in laatste functie: meer dan 1 - 2 jaar	0,76	0,68	0,85	-0,27	0,06	-4,99	< 0,001	
		Aantal jaren in laatste functie: meer dan 2 - 5 jaar	0,78	0,70	0,86	-0,25	0,05	-5,01	< 0,001	
		Aantal jaren in laatste functie: meer dan 5 - 10 jaar	0,56	0,50	0,62	-0,59	0,06	-10,71	< 0,001	
		Aantal jaren in laatste functie: meer dan 10 - 50 jaar	0,46	0,41	0,51	-0,77	0,06	-14,14	< 0,001	
X	7	Visie op terugkeer naar werk	1,39	1,34	1,45	0,33	0,02	15,62	< 0,001	0,014
	30	Bemiddelingsberoep soort: Economisch-administratieve beroepen (ref)	1,00							0,010
		Bemiddelingsberoep soort: Agrarische beroepen	2,04	1,61	2,57	0,71	0,12	5,95	< 0,001	
		Bemiddelingsberoep soort: Informatica beroepen	1,22	1,02	1,46	0,20	0,09	2,13	0,033	
		Bemiddelingsberoep soort: Medische en paramedische beroepen	1,28	1,09	1,50	0,25	0,08	2,95	0,003	
		Bemiddelingsberoep soort: Openbare orde- en veiligheidsberoepen	1,01	0,75	1,36	0,01	0,15	0,06	0,953	
		Bemiddelingsberoep soort: Pedagogische beroepen	1,02	0,84	1,25	0,02	0,10	0,21	0,835	
		Bemiddelingsberoep soort: Sociaal-culturele beroepen	0,99	0,88	1,10	-0,01	0,06	-0,25	0,804	
		Bemiddelingsberoep soort: Technische en industrieberoepen	1,77	1,61	1,93	0,57	0,05	12,19	< 0,001	
		Bemiddelingsberoep soort: Transportberoepen	1,41	1,25	1,60	0,35	0,06	5,47	< 0,001	
		Bemiddelingsberoep soort: Verzorgende en dienstverlenende beroepen	1,16	1,06	1,27	0,15	0,05	3,21	0,001	
	5	Samenstelling huishouden: woont alleen	1,00							0,008
		Samenstelling huishouden: getrouwd/woont samen, geen thuiswonende kinderen	1,22	1,11	1,33	0,20	0,05	4,28	< 0,001	
		Samenstelling huishouden: man getrouwd/woont samen, jongste kind thuis 0-6 jaar	1,25	1,11	1,41	0,22	0,06	3,61	< 0,001	
		Samenstelling huishouden: vrouw getrouwd/woont samen, jongste kind thuis 0-6 jaar	1,03	0,91	1,16	0,03	0,06	0,45	0,655	

WV	Nr	Factor	OR	BI		b	SE	z	p	ΔR <sup>2</sup>
				2,5%	97,5%					
1.0										
		Samenstelling huishouden: getrouwd/woont samen, jongste kind thuis 7 jaar en ouder	1,55	1,42	1,70	0,44	0,05	9,47	< 0,001	
		Samenstelling huishouden: alleenstaande ouder	0,93	0,82	1,05	-0,08	0,06	-1,22	0,222	
		Samenstelling huishouden: anders	1,03	0,90	1,17	0,03	0,07	0,39	0,694	
	35	Maximale duur WW-recht langer dan 12 maanden: nee (ref)	1,00							0,007
		Maximale duur WW-recht langer dan 12 maanden: ja	0,59	0,54	0,65	-0,52	0,05	-11,44	< 0,001	
	29	Bedrijfstak voor WW: (Financiële) dienstverlening (ref)	1,00							0,007
		Bedrijfstak voor WW: Bouw en Hout	1,85	1,60	2,14	0,62	0,08	8,18	< 0,001	
		Bedrijfstak voor WW: Gezondheid	1,17	1,06	1,29	0,16	0,05	3,21	0,001	
		Bedrijfstak voor WW: Industrie sec	1,25	1,12	1,40	0,22	0,06	3,98	< 0,001	
		Bedrijfstak voor WW: Landbouw, Visserij, Voeding	1,38	1,17	1,64	0,32	0,09	3,72	< 0,001	
		Bedrijfstak voor WW: Overig	1,14	0,56	2,31	0,13	0,36	0,35	0,723	
		Bedrijfstak voor WW: Overig bedrijf en beroep	1,24	1,10	1,40	0,22	0,06	3,55	< 0,001	
		Bedrijfstak voor WW: Transport	1,29	1,12	1,49	0,26	0,07	3,60	< 0,001	
		Bedrijfstak voor WW: Uitzendbedrijven	1,62	1,44	1,83	0,48	0,06	8,02	< 0,001	
		Bedrijfstak voor WW: Winkelbedrijf en Groothandel	1,04	0,95	1,14	0,04	0,05	0,77	0,441	
X	8	Werkzoekgedrag m.b.t. direct contact werkgevers	1,27	1,21	1,33	0,24	0,02	9,79	< 0,001	0,005
X	14	Gezondheidsbeleving	1,29	1,20	1,38	0,25	0,04	7,28	< 0,001	0,003
	18	Balans voor- en nadelen niet-werken: evenveel nadelen als voordelen (ref)	1,00							0,002
		Balans voor- en nadelen niet-werken: meer nadelen dan voordelen	1,23	1,15	1,32	0,21	0,04	5,75	< 0,001	
		Balans voor- en nadelen niet-werken: meer voordelen dan nadelen	0,87	0,69	1,10	-0,14	0,12	-1,14	0,253	
	36	WIA 35-min-indicatie: Nee	1,00							0,002
		WIA 35-min-indicatie: Ja	0,50	0,39	0,64	-0,70	0,13	-5,55	< 0,001	
X	6	Problemen Nederlandse Taal: geen problemen (ref)	1,00							0,002
		Problemen Nederlandse Taal: beetje problemen	0,76	0,66	0,88	-0,27	0,07	-3,62	< 0,001	
		Problemen Nederlandse Taal: veel problemen/kan geen NL	0,63	0,50	0,79	-0,47	0,12	-4,03	< 0,001	
	20	Acceptatiebereidheid m.b.t. voltijd baan	1,08	1,05	1,11	0,08	0,02	5,15	< 0,001	0,001

WV	Nr	Factor	OR	BI		b	SE	z	p	$\Delta R^2$
				2,5%	97,5%					
1.0										
X	13	Algemeen werkvermogen	1,07	1,05	1,11	0,07	0,01	5,02	< 0,001	0,001
	17	Aantal uren in staat te werken	1,01	1,01	1,01	0,01	0,00	4,64	< 0,001	0,001
	28	Opleidingsniveau: laag (ref)	1,00							0,001
		Opleidingsniveau: midden	1,19	1,10	1,28	0,17	0,04	4,42	< 0,001	
		Opleidingsniveau: hoog	1,10	1,00	1,21	0,10	0,05	2,03	0,043	
	19	Acceptatiebereidheid m.b.t. afwijkende tijden	1,06	1,03	1,09	0,06	0,01	4,35	< 0,001	0,001
	2	Inkomsten uit werk naast WW: nee (ref)	1,00							0,001
		Inkomsten uit werk naast WW: ja	1,17	1,08	1,27	0,16	0,04	3,82	< 0,001	

### 3.6 Alternatieve methoden

Naast het hierboven besproken logistische regressiemodel hebben we ook de volgende alternatieve regressiemethoden toegepast: penalized regressiemethoden (ridge regressie, lasso regressie en elastic net), support vector machine en random forest. Voor meer inzicht in deze methoden is het raadzaam om eerst bijlage C te lezen. We pasten deze methoden toe op een eerdere versie van het onderzoekbestand<sup>18</sup>. We vergeleken de resultaten van de sensitiviteit, specificiteit en percentage correct geclassificeerden in de testset tussen de verschillende methodes. Aan de hand van het percentage correct geclassificeerden en de som SN + SP concludeerden we dat random forest een iets beter resultaat gaf dan het logistisch regressiemodel: het percentage correct geclassificeerden was 70% ten opzichte van 69% bij het multivariate logistisch regressie model zonder tegengestelde verbanden. De overige alternatieve modellen deden het niet beter. We hebben overwogen om de oplossing van random forest te kiezen als uiteindelijke model. Random forest is echter een black box methode, die wellicht wel iets beter voorspelt, maar zich slecht leent voor interpretatie en implementatie (zie ook bijlage C). Dit laatste is juist in het gebruik van de Werkverkenner een belangrijk aspect. Bij random forest wordt de uiteindelijk voorspelde status werkherhvatting voor een individu namelijk gebaseerd op meerdere classificatiebomen (in ons geval 1.000). Omdat het verschil in voorspelkracht tussen random forest en logistische regressie slechts 1% is, volstaan we met het logistisch regressiemodel.

### 3.7 Logistische regressie met factoren Werkverkenner 1.0

**Conclusie:** De voorspelkracht van het model voor de voorspelling van status werkherhvatting binnen 12 maanden met de 11 factoren van Werkverkenner 1.0 is iets lager (68% correct voorspeld; AUC = 0,75; Brier = 0,20) dan het meest spaarzame logistisch regressie model met 18 factoren (70% correct voorspeld; AUC = 0,78; Brier = 0,19). Globaal kunnen we spreken van een redelijk discriminerend ( $0,70 < AUC < 0,80$ ) model, maar een iets minder goed passend model (Brier = 0,20). Negen van de 11 factoren hebben een significant effect op de voorspelling van status werkherhvatting binnen 12 maanden. Niet significant zijn Problemen Nederlandse taal en Werkzoek-intentie.

We voeren eveneens analyses uit met de elf factoren van Werkverkenner 1.0. De variabelen die meegaan zijn Leeftijd, Aantal jaren gewerkt in laatste functie, Problemen met de Nederlandse taal, Visie op terugkeer naar werk, Werkzoekgedrag m.b.t. direct contact werkgevers, Werkzoekintentie, Externe variabele attributie, Algemeen werkvermogen, Psychisch werkvermogen, Lichamelijk werkvermogen en Te ziek voelen om te werken. We gebruiken dus niet de gewichten van Werkverkenner 1.0, maar schatten de gewichten opnieuw met de ontwikkelset van het huidige onderzoek. De voorspellende eigenschappen van de modellen voor Werkverkenner 1.0 worden bepaald voor de testset. De resultaten

<sup>18</sup> Dit betreft hetzelfde onderzoekbestand als in deze rapportage, maar de onderzoekspopulatie is een kortere periode gevolgd (i.e. het 2e tussenbestand). De status werkherhvatting kon op het moment van het vervaardigen van het onderzoekbestand hooguit negen maanden na aanvang WW worden vastgesteld in plaats van 12 maanden. De percentages correct geclassificeerden hebben in deze paragraaf betrekking op de modellen die kans op werk binnen negen maanden voorspellen die geanalyseerd zijn op het 2<sup>e</sup> tussenbestand.

voor status werkhervatting binnen 12 maanden en binnen 9 maanden zijn weergegeven in, respectievelijk, tabellen 3.7a en 3.7b.

Tabel 3.7a: Voorspellende eigenschappen van het meervoudige logistisch regressie-model met de factoren van Werkverkenner 1.0 voor de voorspelling van **Status werkhervatting binnen 12 maanden**;  $n = 13.269$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
22,5	95,0	0,78	57,5	117,5	212,5	0,751	0,201
35,6	90,0	0,72	61,8	125,6	215,6		
45,1	85,0	0,67	64,4	130,1	215,1		
53,3	80,0	0,63	66,2	133,3	213,3		
59,5	75,0	0,59	67,0	134,5	209,5		
65,2	70,0	0,55	67,5	135,2	205,2		
70,0	65,6	0,52	67,9	135,6	201,2		
<b>75,0</b>	<b>60,6</b>	<b>0,48</b>	<b>68,0</b>	<b>135,6</b>	<b>196,1</b>		
80,0	54,7	0,44	67,8	134,7	189,4		
85,0	48,2	0,40	67,2	133,2	181,3		
90,0	40,3	0,34	66,0	130,3	170,7		
95,0	30,0	0,27	63,7	125,1	155,1		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met de hoogste waarde voor % correct voorspeld heeft een donkerblauwe balk.

Tabel 3.7b: Voorspellende eigenschappen van het meervoudige logistisch regressie-model met de factoren van Werkverkenner 1.0 voor de voorspelling van **Status werkhervatting binnen 9 maanden**;  $n = 13.269$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
21,4	95,0	0,69	63,3	116,4	211,4	0,733	0,205
32,3	90,0	0,64	65,2	122,3	212,3		
41,5	85,0	0,59	66,3	126,5	211,5		
<b>49,9</b>	<b>80,0</b>	<b>0,55</b>	<b>67,1</b>	<b>129,9</b>	<b>209,9</b>		
56,0	75,0	0,51	66,8	131,0	206,0		
62,1	70,0	0,48	66,6	132,1	202,1		
70,0	63,0	0,43	66,0	133,0	196,0		
75,0	57,9	0,39	65,2	132,9	190,8		
80,0	51,7	0,36	63,9	131,7	183,4		
85,0	45,5	0,32	62,5	130,5	176,0		
90,0	37,6	0,27	60,2	127,6	165,2		
95,0	27,0	0,21	56,3	122,1	149,1		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met de hoogste waarde voor % correct voorspeld heeft een donkerblauwe balk.

Over het algemeen zijn de voorspellende eigenschappen iets lager dan de meest spaarzame modellen met respectievelijk 18 factoren (voor 12 maanden; tabel 3.5a) en 17 factoren (voor 9 maanden; tabel 3.5b). De verschillen zijn echter niet groot. Het model met de factoren van Werkverkenner 1.0 voor de voorspelling van status werkhervatting binnen 12 maanden (tabel 3.7a) en het model voor de voorspelling van status werkhervatting binnen



9 maanden (tabel 3.7b) kunnen we globaal beschouwen als redelijk discriminerende ( $0,70 < \text{AUC} < 0,80$ ) modellen, maar iets minder goed passende modellen ( $\text{Brier} \geq 0,20$ ).

De verklarende eigenschap (Nagelkerke  $R^2$ ) is 0,258 voor het model van status werkhervatting binnen 12 maanden en 0,225 voor het model van status werkhervatting binnen 9 maanden.

Bijlage L geeft een overzicht van de invloed per factor van Werkverkenner 1.0 aan de voorspelling van status werkhervatting binnen 12 maanden in het meervoudige logistisch regressiemodel. Negen van de 11 factoren dragen significant bij aan de voorspelling. De effecten van Problemen Nederlandse taal en Werkzoekintentie zijn niet significant (zie bijlage L, tabel L.1). Het effect van Externe variabele attributie is tegengesteld aan het verwachte effect: een *toename* in Externe variabele attributie, dat wil zeggen het toeschrijven van het vinden van werk aan oorzaken buiten jezelf, zorgt voor een *grotere* kans op Werkhervatting. Acht van de negen significante factoren zitten ook in het meest spaarzame, interpreteerbare model met 18 factoren (waarbij de factor Gezondheidsbeleving een schaal is gebaseerd op o.a. drie factoren van Werkverkenner 1.0 en waarbij de factor Externe variabele attributie ontbreekt).

Tien van de 11 factoren van Werkverkenner 1.0 dragen significant bij aan de voorspelling van status werkhervatting binnen 9 maanden. Alleen het effect Werkzoekintentie is niet significant. Ook in dit model is het effect van Externe variabele attributie tegengesteld aan het verwachte effect.

## 4 Resultaten voorspelling werkloosheidsduur

**Algehele conclusie:** Cox regressieanalyses zijn uitgevoerd met als uitkomstvariabele de tijd tot werkhervatting (in dagen), rekening houdend met de status werkhervatting aan het einde van de waarnemingsperiode (i.e. of werkhervatting heeft plaatsgevonden of niet). We hebben in de analyses dezelfde stappen gevolgd als in de meervoudige logistische regressie analyse van hoofdstuk 3.

Het meest spaarzame, interpreteerbare Cox regressiemodel bevat 20 factoren en heeft een redelijk goed onderscheidend vermogen op de test set ( $C\text{-index} = 0,703$ ). Zeventien van de 20 factoren komen overeen met het meest spaarzame, interpreteerbare logistische regressiemodel van hoofdstuk 3. De unieke bijdrage van de drie nieuwe factoren is relatief klein ( $\Delta R^2 \leq 0,001$ ). Het resultaat van het Cox regressiemodel is een survival curve die de kans op werkloos te blijven weergeeft vanaf aanvang van de werkloosheid ( $t = 0$ ) tot en met het einde van de waarnemingsperiode. Op basis hiervan kunnen we de kans op werkhervatting schatten op een bepaald moment (bijv.  $t = 365$ ; dus binnen 12 maanden). Het hoogste percentage correct voorspeld van het Cox regressiemodel (met  $t = 365$ ) is gelijk aan 70,4% en de bijbehorende Brierscore is goed (0,19).

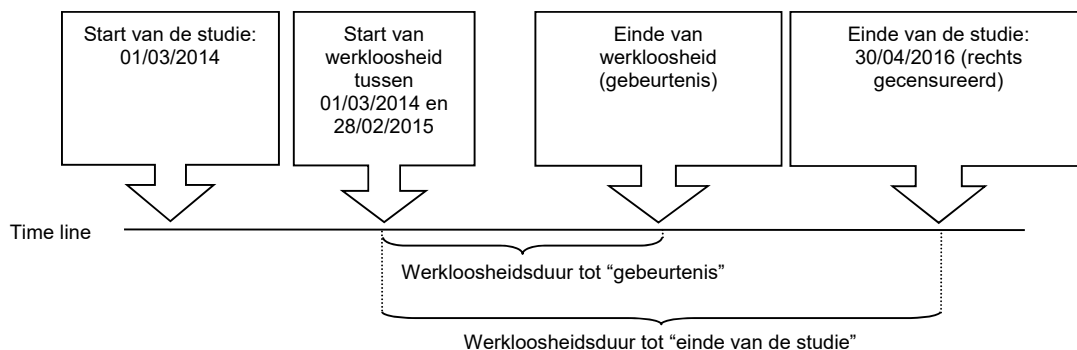
UWV wil ook graag bij aanvang van de WW-uitkering een uitspraak kunnen doen over de verwachte werkloosheidsduur per werkzoekende. Henderson & Keiding (2005) waarschuwen echter voor het doen van individuele 'time to event' schattingen op basis van bijv. een Cox regressiemodel, omdat deze erg onzeker zijn. Ook in ons geval blijkt de voorspelling niet zo nauwkeurig.

*Wat zijn voorspellende persoons- en situatietekenen voor de duur van de werkloosheid van de WW-instroom? en in welke mate kan het model een goede voorspelling geven van de werkloosheidsduur van de WW'er bij aanvang van de WW?* hebben we met behulp van duuranalyse onderzocht. We gebruiken als duurmodel een Cox proportional hazards model (oftewel Cox regressiemodel). Een dergelijke analyse heeft als doel om te onderzoeken of de variatie in de kans op de gebeurtenis, het hervatten van werk, systematisch varieert met de predictoren, de factoren van de Werkverkenner. De Cox regressieanalyse valt onder de familie van *Survivalanalyse*. Dit type analyse bekijkt *of* en *wanneer* een gebeurtenis, werkhervatting, plaatsvindt. Voor het uitvoeren van een Cox regressie analyse zijn twee uitkomstvariabelen belangrijk: het plaatsvinden van de gebeurtenis, dus de status werkhervatting, en de tijd totdat de gebeurtenis plaatsvindt, *time to event*. Zie bijlage C, voor een uitgebreide toelichting.

Voor de duuranalyses hebben we als waarnemingsperiode genomen de periode vanaf aanvang van de werkloosheid, tussen 1 maart 2014 en 28 februari 2015, tot en met het einde van de studie op 30 april 2016. We maken verder gebruik van twee nieuwe uitkomstvariabelen: 1) duur werkloosheid in dagen, vanaf de dag van werkloosheid tot aan de dag van werkhervatting of bij geen werkhervatting tot aan de dag waarop de studie beëindigd is)<sup>19</sup>; 2) status werkhervatting binnen de waarnemingsperiode tot en met 30 april 2016 (zie figuur 4.1).

<sup>19</sup> Het gaat hierbij om de duur in dagen tot de eerste gemeten werkhervatting binnen de waarnemingsperiode (i.e. "gebeurtenis"). Op een later moment (ook binnen de waarnemingsperiode) kan iemand weer werkloos zijn geworden.

Een belangrijk onderscheid tussen de verschillende survivalmodellen is of de duur gemeten wordt op discrete of continue momenten (d.w.z. op ieder moment in de waarnemingsperiode). Wij hebben hier te maken met een duur die continu gemeten is.



Figuur 4.1: Overzicht van de tijdslijn van de analyses voor werkloosheidsduur.

De analyses zijn uitgevoerd op dezelfde ontwikkel-, validatie- en testset als gebruikt in hoofdstuk 3. Ook hebben we dezelfde factoren gebruikt (weergegeven in tabel 1.1). Personen die op een bepaald moment gedurende de waarnemingsperiode overleden of gepensioneerd zijn hebben we niet meegenomen<sup>20</sup>. Vergeleken met het bestand dat we gebruikten in hoofdstukken 2 en 3 betrof dit slechts een zeer kleine groep (0,2%)<sup>21</sup>. De analyses volgen dezelfde 3 stappen als bij de meervoudige logistische regressie analyses: in stap 1 schatten we het duurmodel met alle factoren samen; in stap 2 schatten we het interpreteerbare, meervoudige model (d.w.z. zonder de factoren die moeilijk te interpreteren zijn); in stap 3 passen we modelselectie toe om het interpreteerbare, meest eenvoudige model te schatten.

#### 4.1 Voorspelling van werkloosheidsduur met alle factoren samen

In stap 1 van de analyses schatten we het Cox model met alle factoren samen (zie bijlage G). Dit model bevatte een aantal verbanden die een andere richting lieten zien dan verwacht (zie bijlage G). Het betreft de volgende zes factoren: Werkzoekattitude m.b.t. voordeel en aangenaamheid van werk zoeken, Gemiddeld aantal arbeidsuren voor werkloosheid, Self-efficacy m.b.t. sollicitatievoorbereiding, Werkzoekattitude m.b.t. nut en noodzaak van werk zoeken, Externe variabele attributie en Acceptatiebereidheid m.b.t. onwenselijk werk. De unieke bijdrage van ieder van deze factoren in het model is relatief klein (relatieve Cox en Snell  $R^2 \leq 0,002$ , zie bijlage G). We hebben vervolgens het model zonder deze verbanden geschat (stap 2).

<sup>20</sup> Een gebruikelijke strategie is om "pensioen" en/of "overleden" als een aparte gebeurtenis te beschouwen (naast de gebeurtenis werkhervatting) en een "competing risk model" te schatten. Dit model konden we echter niet schatten, aangezien het aantal personen dat deze gebeurtenis mee maakte te klein was in verhouding tot het aantal variabelen.

<sup>21</sup> Het totale onderzoeksbestand gebruikt in Hoofdstuk 2 en 3 bevat 53.079 personen. Uit dit bestand gingen in totaal 142 (0,2%) personen met pensioen of overleden tijdens de langere waarnemingsperiode die we gebruikten voor de duuranalyses. Dit brengt het totaal aantal personen gebruikt voor de duuranalyses op 52.937, waarvan 26.474 in de ontwikkelset, 13.237 in de validatieset en 13.226 in de testset.

## 4.2 Voorspelling van werkloosheidsduur met alleen interpreteerbare factoren

Het Cox regressiemodel met alleen interpreteerbare factoren resulteerde in vergelijkbare voorspellende eigenschappen als het model met alle factoren (zie tabel 4.1).

Tabel 4.1: Voorspellende eigenschappen van de geschatte Cox modellen

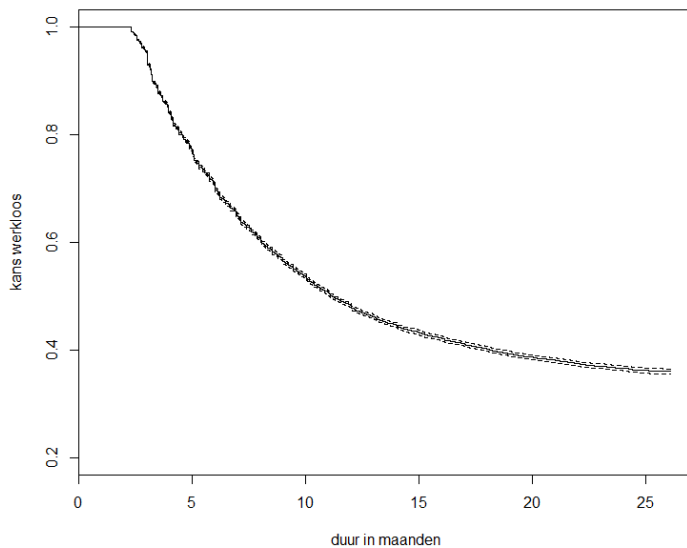
Model	Training set	Test set	
	Cox & Snell $R^2$	C-index <sup>1</sup>	Brierscore <sup>2</sup>
<b>stap 1:</b> Met alle factoren samen	0,279	0,707	0,191
<b>stap 2:</b> Met alleen interpreteerbare factoren	0,276	0,705	0,192
<b>stap 3:</b> Meest spaarzame, interpreteerbare model	0,273	0,704	0,192

<sup>1</sup>C-index is een maat voor het onderscheidend vermogen en is vergelijkbaar met de AUC-index. <sup>2</sup>Het Cox model resulteert in een curve voor de Brierscore over de hele tijdsperiode. De getoonde Brierscore is bepaald op tijdstip  $t = 365$  dagen (12 maanden).

### 4.3 Voorspelling van werkloosheidsduur met meest spaarzame, interpreteerbare model

In een derde stap hebben we stapsgewijze modelselectie toegepast met AIC of BIC en de voorspellende eigenschappen tussen de modellen vergeleken op de validatieset (zie bijlage C voor meer uitleg). De backward BIC methode gaf het beste resultaat op de validatieset. Het resulteerde in een model met 20 factoren. We hebben vervolgens de voorspellende eigenschappen bepaald van dit model op de test set (zie tabel 4.1, stap 3). De C-index en de Brierscore zijn vrijwel gelijk aan de waardes van stap 2.

Het model resulteert o.a. in een zogenaamde geschatte *survivalcurve* voor een individu (en een curve voor de hele groep). Deze curve geeft de geschatte kans om werkloos te blijven (i.e. geen gebeurtenis) weer over de gehele waarnemingsperiode (figuur 4.2).



Figuur 4.2: De geschatte kans om werkloos te blijven ( $y$ -as) voor een specifiek persoon (willekeurig geselecteerd uit het bestand) weergegeven vanaf het moment van werkloosheid (0 maanden) tot en met de maximaal waargenomen duur van werkloosheid in deze studie (26 maanden;  $x$ -as).

Het is mogelijk om de specificiteit en sensitiviteit van een Cox regressiemodel te bekijken voor een gegeven tijdstip  $t$  (Heagerty & Zheng, 2005; zie ook bijlage C). Om de voorspelkracht van het Cox regressiemodel goed te kunnen vergelijken met het logistisch regressiemodel van hoofdstuk 3, hebben wij dit gedaan voor  $t = 365$  dagen (oftewel 12 maanden). De resultaten zijn weergegeven in tabel 4.2. Bij een afkapwaarde van 0,44 bereiken we het hoogste percentage correct geïdentificeerd. Het percentage correct voorspeld van dit model is 70,4 en is dus vrijwel gelijk aan het percentage van het meest spaarzame logistisch regressiemodel (70,3; tabel 3.5a).

Tabel 4.2: Voorspellende eigenschappen van het meest spaarzame, interpreteerbare Cox model voor de voorspelling van **status werkhervatting binnen 12 maanden (i.e.  $t = 365$  dagen)**;  $n = 13.226$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	Brier
99,0	16,9	0,19	59,5	115,9	132,8	0,192
95,0	33,6	0,29	65,5	128,6	162,2	
90,0	45,6	0,36	68,7	135,6	181,2	
85,0	54,0	0,40	70,1	139,0	193,0	
80,0	60,1	0,44	70,4	140,1	200,2	
75,0	64,8	0,47	70,1	139,8	204,6	
70,1	69,2	0,50	69,6	139,3	208,5	
69,1	70,0	0,51	69,5	139,1	209,1	
63,3	74,9	0,55	68,9	138,2	213,1	
57,0	80,0	0,59	68,1	137,0	217,0	
49,6	85,0	0,64	66,6	134,6	219,6	
39,1	90,0	0,70	63,6	129,1	219,1	

*Noot.* Voor deze tabel is de inverse genomen van de kans op werkloosheid (dus  $1 - p(\text{survival})$ ). De regel met de hoogste waarde voor % correct voorspeld is donkerblauw.

De factoren van het meest spaarzame, interpreteerbare Cox regressiemodel zijn weergegeven in tabel 4.3 en geordend naar belangrijkheid op basis van de relatieve toename in  $R^2$ . Alle factoren van het spaarzame, meervoudig logistisch regressiemodel (tabel 3.6) zitten ook in dit Cox regressiemodel met uitzondering van Problemen Nederlandse taal. Als een factor kansvergroterend (of kansverkleinerend) was in het logistisch regressiemodel dan is dit ook het geval in het Cox regressiemodel.

De eerste negen belangrijke factoren zijn gelijk in het Cox regressiemodel en het logistisch regressie model. Er zijn drie factoren in het Cox regressiemodel die niet in het logistisch regressiemodel voorkomen, dit zijn: Nationaliteit, Financiële noodzaak om binnen 2 maanden werk te vinden en Aanvullende uitkering(en) naast WW. De unieke bijdrage van deze drie factoren is echter erg klein (relatieve toename in  $R^2 \leq 0,001$ ; tabel 4.3).

Tabel 4.3: Resultaten voor de voorspelling van werkloosheidsduur met het meest spaarzame, interpreteerbare Cox model (op de ontwikkelset:  $n = 26.474$ ).

WV	Nr	Factor	HR	95% BI		coef	SE	z	p	$\Delta R^2$
				2,5%	97,5%					
1.0										
X	26	Leeftijd	-0,03	0,97	0,97	0,97	0,00	-25,03	< 0,001	0,025
X	7	Visie op terugkeer naar werk	0,23	1,26	1,23	1,29	0,01	19,37	< 0,001	0,014
X	3	Aantal jaren in laatste functie: 0 - 0,5 jaar	0,00	1,00						0,012
		Aantal jaren in laatste functie: meer dan 0,5 - 1 jaar	0,00	1,00	0,95	1,06	0,03	0,00	0,999	
		Aantal jaren in laatste functie: meer dan 1 - 2 jaar	-0,10	0,91	0,86	0,96	0,03	-3,55	< 0,001	
		Aantal jaren in laatste functie: meer dan 2 - 5 jaar	-0,11	0,90	0,85	0,94	0,03	-4,24	< 0,001	
		Aantal jaren in laatste functie: meer dan 5 - 10 jaar	-0,32	0,73	0,69	0,77	0,03	-10,53	< 0,001	
		Aantal jaren in laatste functie: meer dan 10 - 50 jaar	-0,47	0,63	0,59	0,67	0,03	-14,67	< 0,001	
	30	Bemiddelingsberoep soort: Economisch-administratieve beroepen (ref)	0,00	1,00						0,008
		Bemiddelingsberoep soort: Agrarische beroepen	0,40	1,49	1,33	1,67	0,06	6,86	< 0,001	
		Bemiddelingsberoep soort: Informatica beroepen	0,08	1,08	0,97	1,20	0,05	1,46	0,145	
		Bemiddelingsberoep soort: Medische en paramedische beroepen	0,10	1,10	1,00	1,21	0,05	2,02	0,044	
		Bemiddelingsberoep soort: Openbare orde- en veiligheidsberoepen	-0,04	0,96	0,82	1,13	0,08	-0,50	0,620	
		Bemiddelingsberoep soort: Pedagogische beroepen	0,08	1,08	0,97	1,20	0,06	1,36	0,175	
		Bemiddelingsberoep soort: Sociaal-culturele beroepen	0,00	1,00	0,93	1,07	0,03	-0,10	0,924	
		Bemiddelingsberoep soort: Technische en industrieberoepen	0,32	1,38	1,32	1,45	0,02	13,11	< 0,001	
		Bemiddelingsberoep soort: Transportberoepen	0,20	1,22	1,14	1,30	0,03	5,71	< 0,001	
		Bemiddelingsberoep soort: Verzorgende en dienstverlenende beroepen	0,06	1,06	1,01	1,11	0,03	2,17	0,030	
	5	Samenstelling huishouden: woont alleen	0,00	1,00						0,005
		Samenstelling huishouden: getrouwd/woont samen, geen thuiswonende kinderen	0,11	1,12	1,06	1,17	0,03	4,44	< 0,001	
		Samenstelling huishouden: man getrouwd/woont samen, jongste kind thuis 0-6 jaar	0,16	1,17	1,10	1,24	0,03	4,88	< 0,001	
		Samenstelling huishouden: vrouw getrouwd/woont samen, jongste kind thuis 0-6 jaar	0,04	1,04	0,98	1,12	0,03	1,28	0,199	
		Samenstelling huishouden: getrouwd/woont samen, jongste kind thuis 7 jaar en ouder	0,32	1,38	1,31	1,45	0,03	11,94	< 0,001	
		Samenstelling huishouden: alleenstaande ouder	-0,02	0,98	0,91	1,05	0,04	-0,59	0,554	
		Samenstelling huishouden: anders	-0,01	0,99	0,93	1,05	0,03	-0,33	0,741	

WV	Nr	Factor	HR	95% BI		coef	SE	z	p	$\Delta R^2$
				2,5%	97,5%					
1.0										
	29	Bedrijfstak voor WW: (Financiële) dienstverlening (ref)	0,00	1,00						0,004
		Bedrijfstak voor WW: Bouw en Hout	0,41	1,51	1,40	1,63	0,04	10,37	< 0,001	
		Bedrijfstak voor WW: Gezondheid	0,08	1,09	1,03	1,15	0,03	2,89	0,004	
		Bedrijfstak voor WW: Industrie sec	0,12	1,13	1,06	1,20	0,03	3,74	< 0,001	
		Bedrijfstak voor WW: Landbouw, Visserij, Voeding	0,23	1,26	1,15	1,38	0,05	4,80	< 0,001	
		Bedrijfstak voor WW: Overig	0,24	1,27	0,87	1,84	0,19	1,25	0,211	
		Bedrijfstak voor WW: Overig bedrijf en beroep	0,11	1,11	1,04	1,19	0,03	3,17	0,002	
		Bedrijfstak voor WW: Transport	0,15	1,16	1,08	1,26	0,04	3,82	< 0,001	
		Bedrijfstak voor WW: Uitzendbedrijven	0,27	1,30	1,23	1,39	0,03	8,60	< 0,001	
		Bedrijfstak voor WW: Winkelbedrijf en Groothandel	0,02	1,02	0,96	1,07	0,03	0,61	0,541	
X	8	Werkzoekgedrag m.b.t. direct contact werkgevers	0,14	1,15	1,12	1,18	0,01	10,68	< 0,001	0,003
	35	Maximale duur WW-recht langer dan 12 maanden: nee (ref)	0,00	1,00						0,003
		Maximale duur WW-recht langer dan 12 maanden: ja	-0,25	0,78	0,74	0,82	0,02	-10,42	< 0,001	
X	14	Gezondheidsbeleving	0,16	1,18	1,13	1,22	0,02	8,24	< 0,001	0,002
	32	Nationaliteit: 1e nat: NL, 2e nat: geen of Westerse (ref)	0,00	1,00						0,001
		Nationaliteit: 1e of 2e nat: Poolse	-0,22	0,80	0,71	0,91	0,06	-3,43	0,001	
		Nationaliteit: 1e of 2e nat: Niet-Westers (vooral Turks, Marokkaans)	-0,42	0,66	0,56	0,77	0,08	-5,21	< 0,001	
		Nationaliteit: 1e nat: Westers, anders dan NL en Poolse	-0,13	0,88	0,78	0,99	0,06	-2,09	0,036	
	36	WIA 35-min-indicatie: Nee	0,00	1,00						0,001
		WIA 35-min-indicatie: Ja	-0,48	0,62	0,53	0,73	0,08	-5,76	< 0,001	
	18	Balans voor- en nadelen niet-werken: evenveel nadelen als voordelen (ref)	0,00	1,00						0,001
		Balans voor- en nadelen niet-werken: meer nadelen dan voordelen	0,11	1,11	1,06	1,16	0,02	4,76	< 0,001	
		Balans voor- en nadelen niet-werken: meer voordelen dan nadelen	-0,14	0,87	0,76	1,00	0,07	-1,93	0,054	
	20	Acceptatiebereidheid m.b.t. voltijdbaan	0,05	1,05	1,03	1,07	0,01	5,61	< 0,001	0,001
	28	Opleidingsniveau: laag (ref)	0,00	1,00						0,001
		Opleidingsniveau: midden	0,11	1,12	1,07	1,17	0,02	5,21	< 0,001	



WW	Nr	Factor	HR	95% BI		coef	SE	z	p	$\Delta R^2$
				2,5%	97,5%					
1.0										
		Opleidingsniveau: hoog	0,08	1,09	1,03	1,15	0,03	3,16	0,002	
X	13	Algemeen werkvermogen	0,04	1,04	1,03	1,06	0,01	5,15	< 0,001	0,001
	17	Aantal uren in staat te werken	0,01	1,01	1,00	1,01	0,00	5,12	< 0,001	0,001
	22	Financiële noodzaak om < 2 maanden werk te vinden: niet belangrijk (ref)	0,00	1,00						0,001
		Financiële noodzaak om < 2 maanden werk te vinden: belangrijk	0,09	1,09	1,02	1,16	0,03	2,64	0,008	
		Financiële noodzaak om < 2 maanden werk te vinden: Zeer belangrijk	0,15	1,16	1,09	1,24	0,03	4,36	< 0,001	
	19	Acceptatiebereidheid m.b.t. afwijkende tijden	0,03	1,03	1,01	1,04	0,01	3,88	< 0,001	< 0,001
	2	Inkomsten uit werk naast WW: nee (ref)	0,00	1,00						< 0,001
		Inkomsten uit werk naast WW: ja	0,08	1,08	1,04	1,13	0,02	3,54	< 0,001	
	1	Aanvullende uitkering(en) naast WW: nee (ref)	0,00	1,00						< 0,001
		Aanvullende uitkering(en) naast WW: ja	-0,12	0,89	0,83	0,95	0,04	-3,27	0,001	

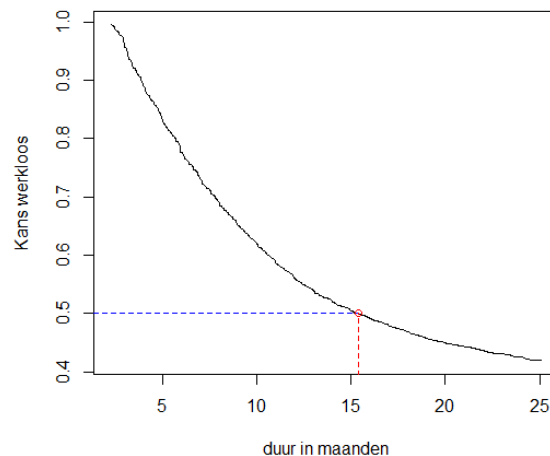
*Noof.* De factoren zijn geordend in volgorde van belangrijkheid op basis van de verklaarde variantie (relatieve toename in Cox & Snell  $R^2$ ). Kansvergrotenende factoren zijn in groen weergegeven, kansverkleinende in rood.

#### 4.4 Schatting van gemiddelde werkloosheidsduur per persoon

**Conclusie:** De schatting van de gemiddelde werkloosheidsduur *per persoon* door het Cox regressiemodel is niet nauwkeurig genoeg. Voor de personen die het werk hervatten tijdens de waarnemingsperiode was de correlatie tussen de geschatte werkloosheidsduur en geobserveerde werkloosheidsduur laag ( $r = 0,18$ ). Dit resultaat bevestigt de resultaten van Henderson & Keiding (2005), die hebben aangetoond dat individuele schattingen van de “time to event”, op basis van bijv. een Cox regressiemodel, erg onzeker zijn.

In deze paragraaf gaan we in op de vraag: *kan het model een betrouwbare voorspelling geven van de werkloosheidsduur van de WW'er bij aanvang van de WW?*

Op basis het meest spaarzame, interpreteerbare Cox regressiemodel kunnen we per persoon een schatting maken van zijn/haar survivalcurve. De survivalcurve kunnen we vervolgens gebruiken om voor een werkzoekende zijn/haar gemiddelde werkloosheidsduur te schatten (i.e. “median survival time”). Deze gemiddelde werkloosheidsduur wordt gewoonlijk gedefinieerd als het tijdstip waarop de kans op werkloosheid voor het individu gelijk is aan 0,50 (= mediaan<sup>22</sup>). Dit is geïllustreerd in figuur 4.3. Voor deze specifieke persoon, is de geschatte gemiddelde werkloosheidsduur gelijk aan 15,4 maanden (468 dagen).



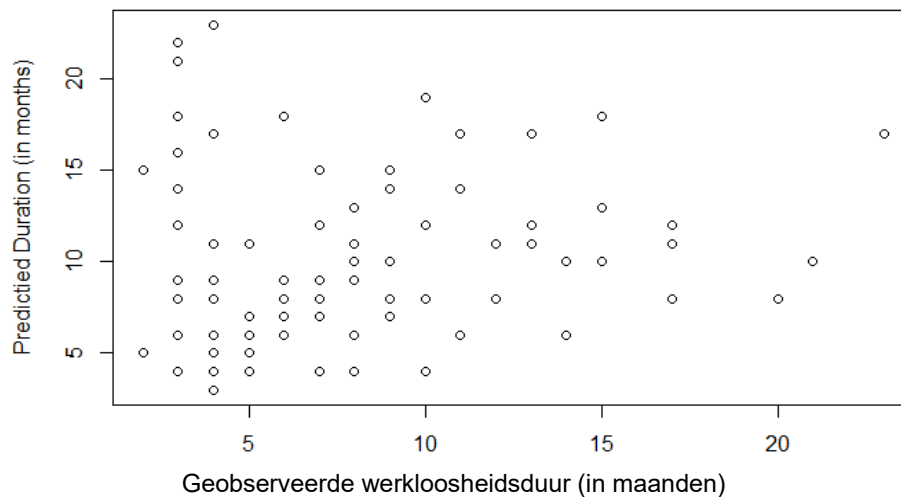
Figuur 4.3: De geschatte gemiddelde werkloosheidsduur (= aangegeven met de blauwe stippellijn), wanneer de kans op werkloosheid gelijk is aan 0,5 (= uitgangspunt van de rode stippellijn).

We hebben voor alle personen in de testset de geschatte gemiddelde werkloosheidsduur op deze manier proberen te bepalen. Echter, voor een redelijk groot aantal personen ( $n = 3.921$ ) was de kans op werkloosheid altijd hoger dan 0,5. Meestal waren dit de personen die nooit het werk hebben hervat. Voor deze personen kon daardoor de geschatte werkloosheidsduur niet bepaald worden. Hun werkloosheidsduur is langer dan de gehele

<sup>22</sup> Het gebruik van de mediaan, of de median survival time, is het meest gebruikelijk voor het schatten van de individuele survival time en tevens gemakkelijk te interpreteren. De kans op werkhervatting na deze periode is gelijk aan 50%. Mogelijk kunnen ook andere waarden gebruikt worden. We verwachten echter dat deze keuze voor de p-waarde weinig invloed heeft op de correlatie tussen de geobserveerde en voorspelde werkloosheidsduur.

waarnemingsperiode van 792 dagen. In de groep personen voor wie de werkloosheidsduur wel bepaald kon worden, hebben we vervolgens de personen geselecteerd met een geobserveerde werkloosheidsduur korter dan 792 dagen. Dat zijn dus de personen die binnen de waarnemingsperiode het werk hervatten. Dit waren 6.975 personen in de testset. Voor hen kunnen we de Pearson correlatie berekenen tussen de geobserveerde werkloosheidsduur en de geschatte gemiddelde werkloosheidsduur. Deze is gelijk aan 0,18 (zie ook figuur 4.4), wat betekent dat het verband zwak is. We concluderen hieruit dat de voorspelling van werkloosheidsduur voor nieuwe werkzoekenden niet nauwkeurig genoeg is.

Dit resultaat bevestigt de resultaten van Henderson & Keiding (2005), die hebben aangetoond dat individuele schattingen van de 'time to event', op basis van bijv. een Cox regressiemodel, erg onzeker zijn.



Figuur 4.4: Verband tussen geobserveerde werkloosheidsduur (x-as) en geschatte werkloosheidsduur (y-as) voor de werkhervatters in de testset ( $n = 6.975$ ).

Het gemiddelde verschil tussen de geschatte werkloosheidsduur en de geobserveerde werkloosheidsduur is gelijk aan 2,2 maanden. Dit betekent dat gemiddeld genomen de werkloosheidsduur met 2,2 maanden overschat wordt.

#### 4.5 Factoren en voorspelmodel Werkverkenner 2.0

Voor UWV is het van belang om te bepalen welke factoren uiteindelijk terecht komen in Werkverkenner 2.0 en welk model gebruikt gaat worden voor de voorspelling van werkhervatting en eventueel werkloosheidsduur. We hebben in dit hoofdstuk gezien dat de voorspellende eigenschappen voor de kans op werkhervatting van het meest spaarzame, interpreteerbare Cox regressiemodel vrijwel gelijk zijn aan die van het meest spaarzame, interpreteerbare logistische regressiemodel. Ook de factoren komen in beide modellen sterk overeen. Tevens hebben we gezien dat de voorspelling van werkloosheidsduur op individueel niveau niet erg nauwkeurig is.

De praktische voordelen van het logistisch regressiemodel zijn 1) het model kan gemakkelijk omgezet worden in een voorspelregel die geïmplementeerd kan worden in de systemen van UWV en vergelijkbaar is met het type voorspelregel van Werkverkenner 1.0 en 2) dat de invloed van de factoren, uitgedrukt als odds ratio, op de kans op werkhervatting iets makkelijker te interpreteren is dan de invloed van de factoren in een Cox regression model, uitgedrukt als hazard ratio. Hieruit vloeit voort dat we het meest spaarzame, interpreteerbare logistisch regressiemodel (gepresenteerd in hoofdstuk 3, tabel 3.6) als beste kandidaat aanraden voor Werkverkenner 2.0.

## 5 Respons en representativiteit

**Conclusie respons op de koplopervestigingen:** In de koplopervestigingen die meededen aan dit onderzoek, verschillen werkzoekenden die de Werkverkenner hebben ingevuld nauwelijks van degenen die de Werkverkenner niet hebben ingevuld. We kunnen concluderen dat de responsgroep geen selecte groep is.

**Conclusie representativiteit koplopervestigingen:** De werkzoekenden die afkomstig zijn van niet-koplopervestigingen verschillen qua achtergrondkenmerken niet van werkzoekenden van de koplopervestigingen. De significante verschillen die we vinden zijn dusdanig klein dat we deze niet relevant achten. We concluderen dat de werkzoekenden van de koplopervestigingen representatief zijn voor alle vestigingen van UWV.

**Conclusie representativiteit geanalyseerde onderzoeksgroep:** De samenstelling van de groep werkzoekenden die de Werkverkenner volledig hebben ingevuld (= de groep die gebruikt is in de analyses) is vergelijkbaar met de samenstelling van de populatie overige werkzoekenden. De significante verschillen die we vinden voor de achtergrondkenmerken zijn dusdanig klein dat we deze niet relevant achten. We concluderen dat de werkzoekenden die de Werkverkenner hebben ingevuld representatief zijn voor alle werkzoekenden.

**Algehele conclusie:** De resultaten die we vinden in dit onderzoek zijn generaliseerbaar naar de landelijke populatie werkzoekenden die voldoen aan de indicatiecriteria voor de Werkverkenner.

In een responsanalyse bekijken we of de groep werkzoekenden die de Werkverkenner heeft ingevuld op demografische kenmerken verschilt van de groep werkzoekenden die de Werkverkenner niet heeft ingevuld, maar wel heeft aangeboden gekregen. We onderzoeken de representativiteit van de werkzoekenden die de Werkverkenner hebben ingevuld.

We onderscheiden voor de respons- en representativiteitsanalyses vier groepen:

- 1) **Werkzoekenden afkomstig van een koploper vestiging die de Werkverkenner hebben ingevuld ( $N = 53.238$ );**
- 2) Werkzoekenden afkomstig van een koplopervestiging die de Werkverkenner aangeboden hebben gekregen maar niet hebben ingevuld ('systematische non-respons':  $N = 23.579$ );
- 3) Werkzoekenden afkomstig van een koplopervestiging die de Werkverkenner niet aangeboden hebben gekregen en dus niet hebben ingevuld ('toevallige non-respons',  $N = 9.750$ ).
- 4) Werkzoekenden die afkomstig zijn van een niet-koploper vestiging ( $N = 201.472$ );

### 5.1 Respons

In een non-responsanalyse bekijken we of de werkzoekenden die de Werkverkenner hebben ingevuld (groep 1) op demografische kenmerken vergelijkbaar zijn met de werkzoekenden die de Werkverkenner wel aangeboden hebben gekregen maar niet hebben ingevuld (groep 2). Dit doen we aan de hand van chi-kwadraattoetsen bij categorische kenmerken en onafhankelijke  $t$ -toetsen bij continue kenmerken. Een 'probleem' bij deze toetsen is dat bij grote groepen kleine verschillen tussen groepen al snel als significant worden bestempeld, terwijl het daadwerkelijke verschil in de praktijk weinig betekenis heeft. Daarom is naast de statistische significantie ook naar effectgroottes gekeken. Net als in hoofdstuk 2 gebruiken we voor categorische variabelen Cramer's  $V$  als maat voor effectgrootte en voor continue variabelen gebruiken we Cohen's  $d$  (i.e. het verschil in gemiddelden uitgedrukt in

standaardafwijkingen). De waarde van Cohen's  $d$  kan zowel positief als negatief zijn. We rapporteren de absolute effect groottes. Voor de richtlijnen voor een klein, middelgroot en groot effect, zie par. 2.3; we ronden hierbij eerst de gevonden waarde af op één decimaal.

### 5.1.1 Resultaten responsanalyse

**Conclusie:** Voor alle onderzochte kenmerken vinden we significante verschillen tussen de werkzoekenden van een koplopervestiging die de Werkverkenner hebben ingevuld en de werkzoekenden van een koplopervestiging die de Werkverkenner niet hebben ingevuld. Het vinden van significante verschillen gebeurt echter snel (en is enigszins vertekenend) vanwege de grote aantallen in de bestanden. Het is daarom beter om naar de effectgroottes te kijken. Alle effectgroottes zijn zeer klein tot klein. Het grootste verschil dat we vinden is dat de werkzoekenden van een koplopervestiging die de Werkverkenner hebben ingevuld gemiddeld iets ouder zijn (42,1 jaar) dan de werkzoekenden van een koplopervestiging die de Werkverkenner niet hebben ingevuld (38,4 jaar; Cohen's  $d = 0,32$ ). We kunnen concluderen dat de respons groep globaal genomen geen selecte groep is.

Tabel 5.1 geeft de resultaten weer van de non-responsanalyse. De werkzoekenden van een koplopervestiging die de Werkverkenner hebben ingevuld zijn gemiddeld ouder (gemiddelde leeftijd 42,1 jaar) dan de werkzoekenden van een koplopervestiging die de Werkverkenner niet hebben ingevuld (gemiddelde leeftijd 38,4 jaar). De gevonden effectgrootte is klein (Cohen's  $d = 0,32$ ). Waarschijnlijk komt het door dit leeftijdsverschil dat de twee groepen ook verschillen op maximale duur recht op WW: de groep die de Werkverkenner heeft ingevuld hebben iets vaker lang recht op WW (46% heeft meer dan 24 maanden recht op WW-uitkering) dan de groep die de Werkverkenner niet hebben ingevuld (31% heeft meer dan 24 maanden recht op WW-uitkering). Ook komt het waarschijnlijk door dit leeftijdsverschil dat het percentage werkhervatting binnen 12 maanden lager is voor de groep die de Werkverkenner heeft ingevuld (52,2%) dan voor de groep die deze niet heeft ingevuld (58,9%). De effectgrootte is echter klein (Cramer's  $V = 0,06$ ). Ook voor de overige variabelen vinden we significante verschillen tussen beide groepen, maar deze zijn zeer klein tot klein ( $0,01 \leq \text{Cramer's } V \leq 0,17$ ).

Tabel 5.1: Resultaten van de non-responsanalyse.

Variabele	Werkzoekenden van een koplopervestiging die de Werkverkenner hebben ingevuld (N = 53.238)	Werkzoekenden van een koplopervestiging die de Werkverkenner niet hebben ingevuld (N = 23.579)	Toetsgegevens en effectgroottes
Werkhervatting binnen 12 mnd (%) <sup>2</sup>	52,2	58,9	$\chi^2(1) = 303,0; p < 0,001; V = 0,06$
Moment van werkhervatting <sup>2</sup>			$\chi^2(3) = 381,8; p < 0,001; V = 0,07$
<i>Geen werk binnen 12 maanden</i>	47,8	41,1	
<i>Werk na 9 en binnen 12 maanden</i>	8,7	8,0	
<i>Werk na 6 en binnen 9 maanden</i>	13,3	16,3	
<i>Werk binnen 6 maanden</i>	30,2	34,6	
Leeftijd (M ± SD)	42,1 (11,6)	38,4 (11,6)	$t(76815) = 40,5; p < 0,001; d = 0,32$
Leeftijd (%)			$\chi^2(3) = 1419,5; p < 0,001; V = 0,14$
<i>t/m 27 jaar</i>	13,0	21,9	
<i>28-49 jaar</i>	56,2	57,1	
<i>50-54 jaar</i>	12,7	9,9	

Variabele	Werkzoekenden van een koploper-vestiging die de Werkverkenner hebben ingevuld (N = 53.238)	Werkzoekenden van een koploper-vestiging die de Werkverkenner niet hebben ingevuld (N = 23.579)	Toetsgegevens en effectgroottes
≥ 55 jaar	18,0	11,1	
Geslacht (% vrouw)	50,9	45,0	$\chi^2(1) = 224,8; p < 0,001; V = 0,05$
Opleiding (%)			$\chi^2(2) = 835,6; p < 0,001; V = 0,10^1$
Laag	20,0	27,6	
Midden	51,3	51,3	
Hoog	28,5	20,4	
Onbekend	0,2	0,6	
Bedrijfstak voor WW (%)			$\chi^2(9) = 1062,3; p < 0,001; V = 0,12$
Landbouw, Visserij, Voeding	3,5	3,4	
Bouw en Hout	5,5	5,2	
Industrie sec	10,3	8,5	
Winkelbedrijf en Groothandel	16,5	14,9	
Transport	5,7	7,4	
(Financ.) Dienstverlening	21,0	17,0	
Uitzend	10,0	16,8	
Gezondheid	19,0	16,1	
Overig bedrijf en beroep	8,5	10,6	
Overig	0,1	0,2	
Bemiddelingsberoep soort			$\chi^2(9) = 947,6; p < 0,001; V = 0,11^1$
Economisch-administratieve beroepen	34,9	26,6	
Agrarische beroepen	1,8	2,2	
Informatica beroepen	2,4	2,0	
Medische en paramedische beroepen	3,3	2,6	
Openbare orde- en veiligheidsberoepen	0,9	1,3	
Pedagogische beroepen	2,1	2,1	
Sociaal-culturele beroepen	7,8	5,5	
Technische en industrieberoepen	20,1	25,3	
Transportberoepen	7,5	10,0	
Verzorgende en dienstverlenende beroepen	19,1	21,8	
Onbekend	0,2	0,6	
Bemiddelingsberoep niveau			$\chi^2(4) = 762,4; p < 0,001; V = 0,10^1$
Elementaire beroepen	8,3	12,9	
Lagere beroepen	31,0	34,9	
Middelbare beroepen	36,9	34,1	
Hogere beroepen	20,1	14,8	
Wetenschappelijke beroepen	3,5	2,6	
Onbekend	0,2	0,6	

Variabele	Werkzoekenden van een koploper-vestiging die de Werkverkenner hebben ingevuld (N = 53.238)	Werkzoekenden van een koploper-vestiging die de Werkverkenner niet hebben ingevuld (N = 23.579)	Toetsgegevens en effectgroottes
Koploper vestiging			$\chi^2(10) = 211,6; p < 0,001; V = 0,05$
<i>Almere (district: Utrecht en Flevoland)</i>	7,3	7,0	
<i>Amsterdam Oost (district: Groot Amsterdam)</i>	7,6	10,5	
<i>Arnhem (district: Gelderland Midden en Zuid)</i>	9,4	9,2	
<i>Breda (district: West Brabant en Zeeland)</i>	15,7	14,9	
<i>Dordrecht (district: Rijnmond)</i>	7,2	7,2	
<i>Haarlem (district Noord Holland Noord)</i>	7,8	6,7	
<i>Midden Holland (district: Den Haag en Leiden)</i>	15,6	14,8	
<i>Leeuwarden (district: Noord)</i>	3,4	3,2	
<i>Zwolle (district: Overijssel en Gelderland Noord)</i>	10,9	11,6	
<i>Sittard (district: Limburg)</i>	4,9	4,5	
<i>Tilburg (district: Midden- en Oost-Brabant)</i>	10,2	10,4	
WIA 35-min (%)	2,0	2,7	$\chi^2(1) = 37,2; p < 0,001; V = 0,02$
Nationaliteit			$\chi^2(3) = 504,7; p < 0,001; V = 0,08^1$
<i>Nederlandse 1e nationaliteit en geen 2e nationaliteit of een westerse 2e nationaliteit (muv Pools)</i>	95,6	91,9	
<i>Poolse 1e of 2e nationaliteit</i>	1,3	3,1	
<i>Niet-westerse 1e of 2e nationaliteit (grotendeels Turks of Marokkaans)</i>	1,1	2,5	
<i>westerse 1e nationaliteit anders dan Nederlandse</i>	1,8	2,1	
<i>Onbekend</i>	0,1	0,3	
Gemiddeld aantal arbeidsuren voor werkloosheid			$\chi^2(3) = 50,7; p < 0,001; V = 0,03$
<i>&lt; 12 uren per week</i>	2,1	2,8	
<i>tussen 12 en 24 uren per week</i>	20,5	19,8	
<i>tussen 25 en 32 uren per week</i>	20,6	20,3	
<i>meer dan 32 uren per week</i>	56,8	57,1	
Maximum duur van de uitkering			$\chi^2(6) = 2237,1; p < 0,001; V = 0,17$
<i>t/m 6 maanden</i>	8,0	15,1	
<i>meer dan 6 t/m 9 maanden</i>	8,3	12,7	



Variabele	Werkzoekenden van een koploper-vestiging die de Werkverkenner hebben ingevuld (N = 53.238)	Werkzoekenden van een koploper-vestiging die de Werkverkenner niet hebben ingevuld (N = 23.579)	Toetsgegevens en effectgroottes
<i>meer dan 9 t/m 12 maanden</i>	8,1	10,7	
<i>meer dan 12 t/m 18 maanden</i>	15,8	17,2	
<i>meer dan 18 t/m 24 maanden</i>	14,3	13,2	
<i>meer dan 24 t/m 36 maanden</i>	29,6	22,5	
<i>meer dan 36 maanden</i>	15,9	8,6	

<sup>1</sup>De categorie 'onbekend' is niet meegenomen in de toets.

<sup>2</sup>De percentages wijken licht af van de percentages werkhervatters die worden genoemd in tabel 3.1, omdat in de eerste kolom van de huidige tabel gebruik is gemaakt van het totale bestand (n = 53.238) in plaats van het bestand zonder missende waarden (N = 53.079).

M = gemiddelde, SD = standaard deviatie; dikgedrukt duidt op een significant toets resultaat ( $p < 0,05$ ).

## 5.2 Representativiteitsanalyse

In de representativiteitsanalyse kijken we op twee manieren naar de *representativiteit* van de werkzoekenden die de Werkverkenner hebben ingevuld. We onderzoeken 1) de representativiteit van de koplopervestigingen en 2) de representativiteit van de geanalyseerde onderzoekspopulatie.

Voor het eerste punt bekijken we of de kenmerken van de werkzoekenden die afkomstig zijn van niet-koplopervestigingen (groep 4) te vergelijken zijn met werkzoekenden die afkomstig zijn van koplopervestigingen (groep 1, 2 en 3). Voor het tweede punt onderzoeken we of de werkzoekenden van de koplopervestigingen die de Werkverkenner hebben ingevuld (groep 1) te vergelijken zijn met de overige werkzoekenden (groep 2, 3 en 4).

### 5.2.1 Resultaten representativiteitsanalyse van de koplopervestigingen

**Conclusie:** De werkzoekenden die afkomstig zijn van niet-koplopervestigingen verschillen qua achtergrondkenmerken niet noemenswaardig van werkzoekenden van de koplopervestigingen. De significante verschillen die we vinden zijn dusdanig klein dat we deze niet relevant achten. We concluderen daarom dat de werkzoekenden van de koplopervestigingen representatief zijn voor de werkzoekenden van alle vestigingen.

De werkzoekenden afkomstig van de niet-koplopervestigingen (groep 4) zijn vergeleken met alle werkzoekenden van de koplopervestigingen (ongeacht of de Werkverkenner is aangeboden of ingevuld: groep 1, 2 en 3). De resultaten staan in tabel 5.2. Er zijn geen verschillen tussen de twee groepen wat betreft status van werkhervatting binnen 12 maanden, gemiddelde leeftijd, gemiddelde aantal arbeidsuren voor werkloosheid en maximale duur van de uitkering. Voor de overige variabelen vinden we een significant toetsresultaat, maar de effectgroottes zijn verwaarloosbaar klein (Cramer's  $V \leq 0,08$ ).

Tabel 5.2: Representativiteit van de koplopervestigingen.

Variabele	Werkzoekenden van een koplopervestiging (N = 86.567)	Werkzoekenden van een niet-koplopervestiging (N = 201.472)	Toetsgegevens en effectgroottes
Werkhervatting binnen 12 mnd (%)	53,7	53,5	$X^2(1) = 0,5; p > 0,05; V = 0,00$
Moment van werkhervatting			$X^2(3) = 8,8; p = 0,03; V = 0,01$
<i>Geen werk binnen 12 Maanden</i>	46,3	46,5	
<i>Werk na 9 en binnen 12 maanden</i>	8,5	8,8	
<i>Werk na 6 en binnen 9 maanden</i>	14,0	13,9	
<i>Werk binnen 6 maanden</i>	31,2	30,9	
Leeftijd (M ± SD)	41,1 (11,7)	41,1 (11,7)	$t(163.272) = 1,30; p > 0,05; d = 0,00$
Leeftijd (%)			$X^2(3) = 6,9; p > 0,05; V = 0,01$
<i>t/m 27 jaar</i>	15,8	15,7	
<i>28-49 jaar</i>	55,8	56,2	
<i>50-54 jaar</i>	12,2	12,2	
<i>≥ 55 jaar</i>	16,2	15,9	
Geslacht (% vrouw)	49,1	49,7	$X^2(1) = 9,3; p < 0,01; V = 0,01$
Opleiding (%)			$X^2(2) = 31,2; p < 0,001; V = 0,01^1$
<i>Laag</i>	22,7	23,1	
<i>Midden</i>	51,5	50,4	
<i>Hoog</i>	25,4	26,2	
<i>Onbekend</i>	0,4	0,3	
Bedrijfstak voor WW (%)			$X^2(9) = 181,6; p < 0,001; V = 0,03$
<i>Landbouw, Visserij, Voeding</i>	3,4	3,4	
<i>Bouw en Hout</i>	5,3	4,8	
<i>Industrie sec</i>	9,5	9,5	
<i>Winkelbedrijf en Groothandel</i>	15,8	16,6	
<i>Transport</i>	6,4	5,6	
<i>(Financ.) Dienstverlening</i>	19,3	20,5	
<i>Uitzend</i>	12,6	12,2	
<i>Gezondheid</i>	18,1	18,0	
<i>Overig bedrijf en beroep</i>	9,4	9,1	
<i>Overig</i>	0,2	0,2	
Bemiddelingsberoep soort			$X^2(9) = 153,1; p < 0,001; V = 0,02^1$
<i>Economisch-administratieve beroepen</i>	31,6	32,7	
<i>Agrarische beroepen</i>	1,9	2,2	
<i>Informatica beroepen</i>	2,2	2,5	
<i>Medische en paramedische beroepen</i>	3,0	3,0	
<i>Openbare orde- en veiligheidsberoepen</i>	1,1	1,1	
<i>Pedagogische beroepen</i>	2,2	2,2	
<i>Sociaal-culturele beroepen</i>	7,0	6,9	
<i>Technische en industrieberoepen</i>	21,9	20,9	

Variabele	Werkzoekenden van een koplopervestiging (N = 86.567)	Werkzoekenden van een niet-koplopervestiging (N = 201.472)	Toetsgegevens en effectgroottes
<i>Transportberoepen</i>	8,6	7,9	
<i>Verzorgende en dienstverlenende beroepen</i>	20,2	20,2	
<i>Onbekend</i>	0,3	0,3	
Bemiddelingsberoep niveau			$\chi^2(4) = 78,5; p < 0,001; V = 0,02^1$
<i>Elementaire beroepen</i>	10,0	9,9	
<i>Lagere beroepen</i>	32,8	31,5	
<i>Middelbare beroepen</i>	35,7	36,0	
<i>Hogere beroepen</i>	18,0	18,7	
<i>Wetenschappelijke beroepen</i>	3,2	3,6	
<i>Onbekend</i>	0,3	0,3	
WIA 35-min (%)	2,3	2,1	$\chi^2(1) = 12,9; p < 0,001; V = 0,01$
Nationaliteit			$\chi^2(3) = 94,6; p < 0,001; V = 0,02^1$
<i>Nederlandse 1e nationaliteit en geen 2e nationaliteit of een westerse 2e nationaliteit (muv Pools)</i>	94,5	93,5	
<i>Poolse 1e of 2e nationaliteit</i>	1,9	2,2	
<i>Niet-westerse 1e of 2e nationaliteit (grotendeels Turks of Marokkaans)</i>	1,5	1,8	
<i>Westerse 1e nationaliteit anders dan Nederlandse</i>	1,9	2,3	
<i>Onbekend</i>	0,2	0,2	
Gemiddeld aantal arbeidsuren voor werkloosheid			$\chi^2(3) = 1,6; p > 0,05; V = 0,00$
<i>&lt; 12 uren per week</i>	2,4	2,3	
<i>tussen 12 en 24 uren per week</i>	20,3	20,4	
<i>tussen 25 en 32 uren per week</i>	20,7	20,7	
<i>meer dan 32 uren per week</i>	56,6	56,6	
Maximum duur van de uitkering			$\chi^2(6) = 11,3; p > 0,05; V = 0,01$
<i>t/m 6 maanden</i>	10,9	11,3	
<i>meer dan 6 t/m 9 maanden</i>	9,3	9,2	
<i>meer dan 9 t/m 12 maanden</i>	8,7	8,7	
<i>meer dan 12 t/m 18 maanden</i>	15,8	15,9	
<i>meer dan 18 t/m 24 maanden</i>	13,6	13,5	
<i>meer dan 24 t/m 36 maanden</i>	27,5	27,4	
<i>meer dan 36 maanden</i>	14,1	13,9	

<sup>1</sup> De categorie 'onbekend' is niet meegenomen in de toets. M = gemiddelde; SD = standaard deviatie; dikgedrukt duidt op een significant toetsresultaat ( $p < 0,05$ ).

## 5.2.2 Resultaten representativiteitsanalyse van de geanalyseerde onderzoeksgroep

**Conclusie:** De samenstelling van de groep werkzoekenden die de Werkverkenner volledig hebben ingevuld is vergelijkbaar met de samenstelling van de populatie overige werkzoekenden. De significante verschillen die we vinden voor de achtergrondkenmerken zijn dusdanig klein dat we deze niet relevant achten. We concluderen daarom dat de werkzoekenden die de Werkverkenner hebben ingevuld representatief zijn voor alle werkzoekenden.

Voor de regressieanalyses selecteren we de werkzoekenden die géén missende waarden hebben op de variabelen die meegaan in de analyses. Tijdens de analyses worden werkzoekenden die een missende waarde op een of meer variabelen hebben namelijk buiten beschouwing gelaten. Van de 53.238 werkzoekenden hebben 53.079 werkzoekenden informatie op alle variabelen die meegaan in de analyses (zie ook par. 3.3). Om te onderzoeken of de werkzoekenden die de Werkverkenner volledig hebben ingevuld en afkomstig zijn van een koplopervestiging ( $N = 53.079$ ) te vergelijken zijn met werkzoekenden die niet gebruikt zijn in de analyses (werkzoekenden die afkomstig zijn van een niet-koploper vestiging ( $N = 201.472$ ), aangevuld met werkzoekenden van koplopervestigingen met systematische ( $N = 23.579$ ) of toevallige non-respons ( $N = 9.750$ ) of missende waarden met name in de administratieve gegevens ( $N = 159$ ), zijn deze twee groepen met elkaar vergeleken op demografische kenmerken. De resultaten staan in tabel 5.3. Het percentage werkzoekenden met een WIA 35-min indicering is in beide groepen gelijk. Er zijn significante verschillen gevonden voor de overige kenmerken, maar de effectgroottes zijn verwaarloosbaar klein (Cramer's  $V \leq 0,06$ , Cohen's  $d \leq 0,10$ ).

Tabel 5.3: Representativiteit van de geanalyseerde onderzoekspopulatie.

Variabele	Werkzoekenden van een koplopervestiging die de Werkverkenner volledig hebben ingevuld ( $N = 53.079$ )	Werkzoekenden van een niet-koplopervestiging en werkzoekenden van een koplopervestiging die de Werkverkenner niet hebben ingevuld ( $N = 234.960$ )	Toets gegevens en effectgroottes
Werkhervatting binnen 12 mnd (%) <sup>2</sup>	52,1	53,9	$\chi^2(1) = 55,4; p < 0,001; V = 0,01$
Moment van werkhervatting <sup>2</sup>			$\chi^2(3) = 65,6; p < 0,001; V = 0,02$
<i>Geen werk binnen 12 Maanden</i>	47,9	46,1	
<i>Werk na 9 en binnen 12 maanden</i>	8,7	8,7	
<i>Werk na 6 en binnen 9 maanden</i>	13,3	14,1	
<i>Werk binnen 6 maanden</i>	30,1	31,2	
Leeftijd (M $\pm$ SD)	42,1 (11,6)	40,9 (11,7)	$t(79.735) = 22,9; p < 0,001; d = 0,10$
Leeftijd (%)			$\chi^2(3) = 486,4; p < 0,001; V = 0,04$
<i>t/m 27 jaar</i>	13,0	16,3	
<i>28-49 jaar</i>	56,2	56,0	
<i>50-54 jaar</i>	12,8	12,1	
<i><math>\geq 55</math> jaar</i>	18,0	15,6	
Geslacht (% vrouw)	50,9	49,2	$\chi^2(1) = 52,4; p < 0,001; V = 0,01$
Opleiding (%)			$\chi^2(2) = 423,0; p < 0,001; V = 0,04^1$

Variabele	Werkzoekenden van een koploper-vestiging die de Werkverkenner volledig hebben ingevuld (N = 53.079)	Werkzoekenden van een niet-koploper-vestiging en werkzoekenden van een koploper-vestiging die de Werkverkenner niet hebben ingevuld (N = 234.960)	Toets gegevens en effectgroottes
<i>Laag</i>	20,0	23,6	
<i>Midden</i>	51,4	50,6	
<i>Hoog</i>	28,6	25,3	
<i>Onbekend</i>			
Bedrijfstak voor WW (%)			$\chi^2(9) = 392,5; p < 0,001; V = 0,04$
<i>Landbouw, Visserij, Voeding</i>	3,5	3,4	
<i>Bouw en Hout</i>	5,5	4,8	
<i>Industrie sec</i>	10,3	9,4	
<i>Winkelbedrijf en Groothandel</i>	16,5	16,3	
<i>Transport</i>	5,7	5,9	
<i>(Financ.) Dienstverlening</i>	21,0	20,0	
<i>Uitzend</i>	10,0	12,8	
<i>Gezondheid</i>	19,0	17,8	
<i>Overig bedrijf en beroep</i>	8,5	9,4	
<i>Overig</i>	0,1	0,2	
Bemiddelingsberoep soort			$\chi^2(9) = 392,9; p < 0,001; V = 0,04^1$
<i>Economisch-administratieve beroepen</i>	35,0	31,8	
<i>Agrarische beroepen</i>	1,8	2,2	
<i>Informatica beroepen</i>	2,4	2,4	
<i>Medische en paramedische beroepen</i>	3,3	3,0	
<i>Openbare orde- en veiligheidsberoepen</i>	,9	1,2	
<i>Pedagogische beroepen</i>	2,1	2,2	
<i>Sociaal-culturele beroepen</i>	7,8	6,7	
<i>Technische en industrieberoepen</i>	20,1	21,5	
<i>Transportberoepen</i>	7,4	8,3	
<i>Verzorgende en dienstverlenende beroepen</i>	19,1	20,5	
<i>Onbekend</i>	0,0	0,4	
Bemiddelingsberoep niveau			$\chi^2(4) = 307,8; p < 0,001; V = 0,03^1$
<i>Elementaire beroepen</i>	8,3	10,3	
<i>Middelbare beroepen</i>	31,0	32,1	
<i>Lagere beroepen</i>	36,9	35,7	
<i>Hogere beroepen</i>	20,2	18,1	
<i>Wetenschappelijke beroepen</i>	3,5	3,4	
<i>Onbekend</i>	0,0	0,4	
WIA 35-min-indicatie (%)	2,0	2,1	$\chi^2(1) = 2,9; p > 0,05; V = 0,00$
Nationaliteit			$\chi^2(3) = 399,7; p < 0,001; V = 0,04^1$
<i>Nederlandse 1e nationaliteit en geen 2e nationaliteit of een westerse 2e nationaliteit</i>	95,7	93,4	

Variabele	Werkzoekenden van een koploper-vestiging die de Werkverkenner volledig hebben ingevuld (N = 53.079)	Werkzoekenden van een niet-koploper-vestiging en werkzoekenden van een koploper-vestiging die de Werkverkenner niet hebben ingevuld (N = 234.960)	Toets gegevens en effectgroottes
<i>(muv Pools</i>			
<i>Poolse 1e of 2e nationaliteit</i>	1,3	2,3	
<i>Niet-westerse 1e of 2e nationaliteit (grotendeels Turks of Marokkaans)</i>	1,1	1,9	
<i>Westerse 1e nationaliteit anders dan Nederlandse</i>	1,8	2,2	
<i>Onbekend</i>	0,0	0,2	
Gemiddeld aantal uren voor werkloosheid			$\chi^2(3) = 19,5; p < 0,001; V = 0,01$
<i>&lt; 12 uren per week</i>	2,1	2,4	
<i>tussen 12 en 24 uren per week</i>	20,5	20,3	
<i>tussen 25 en 32 uren per week</i>	20,6	20,7	
<i>meer dan 32 uren per week</i>	56,8	56,6	
Maximum duur van de uitkering			$\chi^2(6) = 1.032,7; p < 0,001; V = 0,06$
<i>t/m 6 maanden</i>	7,9	11,9	
<i>meer dan 6 t/m 9 maanden</i>	8,3	9,4	
<i>meer dan 9 t/m 12 maanden</i>	8,1	8,9	
<i>meer dan 12 t/m 18 maanden</i>	15,8	15,9	
<i>meer dan 18 t/m 24 maanden</i>	14,3	13,4	
<i>meer dan 24 t/m 36 maanden</i>	29,7	26,9	
<i>meer dan 36 maanden</i>	15,9	13,6	

<sup>1</sup>De categorie 'onbekend' is niet meegenomen in de toets.

<sup>2</sup> De percentages werkhervatters en niet-werkhervatters in de eerste kolom van deze tabel komen overeen met tabel 3.1 (totale bestand).

M = gemiddelde; SD = standaard deviatie; dikgedrukt duidt op een significant toetsresultaat ( $p < 0,05$ ).

## 6 Diagnose: normering en gebruik

**Conclusie normering:** Voor het presenteren van de uitslag van een werkzoekende op Werkverkenner 2.0 aan de adviseurs werk maken we gebruik van een indeling in vijf kleuren, van rood naar donkergroen. De kleurenschakering geeft per factor de interpretatie van de score weer, variërend van belemmerend naar bevorderend voor het hervatten van werk binnen 12 maanden. Voor numerieke factoren gebruiken we als normgroep de groep werkhervatters binnen 12 maanden en bepalen we de vijf kleuren op basis van het gemiddelde en de standaarddeviatie in deze groep met behulp van afkapwaarden op de T-schaal. Voor categorische factoren werken we direct met het criterium status werkhervatting en bepalen we de indeling in kleuren op basis van de odds ratio's in het logistische regressiemodel met 18 factoren (ontwikkeld in hoofdstuk 3). Op basis van de uiteindelijke visualisatie van de scores van een werkzoekende op de 18 factoren kunnen de adviseurs werk snel een diagnose maken. Het is verder mogelijk om de toename in kans op werkhervatting te laten zien als de score op een factor verbetert.

**Conclusie gebruik:** De werkzoekenden die deelnamen aan een kwalitatief gebruikersonderzoek waren over het algemeen positief over de vragenlijst, de gebruiksvriendelijkheid en over de vragen zelf. Alle deelnemers, behalve één deelnemer op één onderdeel, beoordeelden alle onderdelen met een voldoende of een goed. Op basis van de resultaten kunnen we concluderen dat de Werkverkenner door de deelnemers als gebruiksvriendelijk wordt ervaren. Wel kwam voren dat het voor deelnemers soms niet duidelijk is waarom zij de Werkverkenner moeten invullen, wat het nut voor henzelf is en wat ze kunnen verwachten van UWV. Ook kwam naar voren dat de deelnemers behoefte hebben aan een rapportage die beter aansluit bij hun specifieke situatie (meer maatwerk). Het merendeel van de deelnemers ervaart de informatie als reeds bekend en te algemeen.

De adviseurs van UWV die deelnamen aan een kwalitatief gebruikersonderzoek vonden een aantal toekomstscenario's voor terugkoppeling van het profiel van de Werkverkenner van een individuele klant over het algemeen goed. Zij hebben een voorkeur voor een meer genuanceerde presentatie, meer dan de huidige 2 kleuren rood en groen. De weergave is echter nog niet compact genoeg en de tekst eromheen moet flink worden gereduceerd. Verder is er behoefte aan mogelijkheden tot het bekijken van extra resultaten en informatie op verzoek, bijv. door te klikken op een i-button en er is behoefte aan handvaten voor acties naar aanleiding van de resultaten, bijv. voorbeeldvragen of verwijzing naar e-producten.

Dit hoofdstuk gaat over het stellen van een eerste diagnose op basis van de scores op de factoren van de Werkverkenner. We gaan in op de normering van de 18 voorspellende factoren van Werkverkenner 2.0 (par. 6.1 t/m 6.3) en op het gebruik van de Werkverkenner in de praktijk door klanten en adviseurs werk van UWV (par. 6.4). De volgende onderzoeksvragen komen aan bod m.b.t. de normering:

1. Op welke wijze kan vastgesteld worden of de voorspellers van werkhervatting, opgenomen in Werkverkenner 2.0, belemmerend of bevorderend werken voor werkhervatting van de individuele klant?
2. Wat is het effect per voorspeller op de kans op werkhervatting als deze voorspeller gunstig wordt beïnvloed?

En de volgende vragen worden beantwoord m.b.t. het gebruik van de Werkverkenner:

3. Zijn de instructies duidelijk voor de WW-klienten?
4. Zijn de vragen begrijpelijk voor de WW-klienten?
5. Is het invullen van de Werkverkenner gebruiksvriendelijk?
6. Is de rapportage na invulling van de vragenlijst voor zowel de WW-klant als de adviseurs werk klantvriendelijk?

## 6.1 Normering

### 6.1.1 Normeringsmethode

Voor de terugkoppeling van de scores van een nieuwe WW-klant op de factoren van Werkverkenner 2.0 aan de adviseurs werk bij UWV is het wenselijk dat deze eenvoudig is en gemakkelijk te interpreteren. Adviseurs werk hebben als taak de werkzoekende te begeleiden in het vinden van werk. Bij het uitvoeren van deze taak is het nodig om inzicht te hebben welke factoren van een individuele klant mogelijk belemmerend zijn voor het hervatten van werk. Een lijst van de scores op de 18 factoren is lastig omdat de betekenis achter de score niet meteen duidelijk is. De schaal van de factoren is immers niet gelijk en de mate waarin iedere factor aan de voorspelling van werkhervatting bijdraagt verschilt.

In dit hoofdstuk streven we naar het omzetten van de scores op de factoren van Werkverkenner 2.0 naar *normscores* die de adviseur snel een beeld geven van de mate van belemmering voor werkhervatting. We willen de adviseur werk inzicht geven in de betekenis van de scores van een nieuwe WW-klant t.o.v. de WW-populatie die het werk heeft hervat binnen een jaar. De groep werkhervatters binnen 12 maanden ( $n = 27.670$ ) uit het totale onderzoekbestand ( $N = 53.079$ ) gebruiken we daarom als normgroep om te bepalen of een klant belemmerend scoort op bepaalde factoren.

#### 6.1.1.1 Keuze van methode en aantal categorieën van de normschaal

In Werkverkenner 1.0 wordt bij de terugkoppeling aan de adviseurs werk uitsluitend aangegeven of een factor belemmerend is, rood, of niet, groen. De score op een factor wordt voor hen dus weergegeven in twee categorieën. Uit een recente focusgroep met adviseurs werk (zie par. 6.2) bleek dat de adviseurs de voorkeur geven aan een weergave in meer dan twee kleuren, namelijk een indeling in vier of vijf kleuren, van donker rood naar donker groen, van belemmerend naar bevorderend. Een indeling in meer dan vijf categorieën vinden wij niet bruikbaar voor de praktijk, omdat aan iedere categorie een handelingsoptie gekoppeld moet worden voor de adviseurs. We werken daarom in dit hoofdstuk zoveel mogelijk met een indeling van de scores per factor in vijf categorieën.

In de literatuur zijn verschillende methoden voor handen om normen te bepalen. Deze methoden zijn alle van toepassing op *numerieke* factoren (i.e. factoren gemeten op een interval- of ratioschaal). Veelgebruikte normeringsschalen in de sociale wetenschappen zijn 1) de centiel- of percentielschaal, 2) de stanine en 3) de T-schaal (lineair<sup>23</sup>; Eggen & Sanders, 1993). We hebben deze drie normeringsschalen berekend voor de numerieke factoren van Werkverkenner 2.0 met gebruikmaking van de groep werkhervatters binnen 12

---

<sup>23</sup> Er zijn twee manieren om een score om te zetten naar een T-score: de lineaire T-score en de genormaliseerde T-score. Omdat het resultaat van de genormaliseerde T-score omgezet naar vijf categorieën niet veel verschilde van de staninemethode, hebben we ons beperkt tot de lineaire T-score.



maanden als normgroep. Na berekening van de volledige normeringsschaal, bijv. voor de percentielen van 1 tot 100, brachten we de schalen terug naar 5 categorieën.

Voor de categorische factoren van de Werkverkenner 2.0 hebben we een andere benaderingswijze gekozen. Het einddoel is echter hetzelfde als voor de numerieke factoren: we willen de score op de factor uitdrukken in belemmerend of bevorderend voor de kans op werkhervatting. We streven ernaar om dit ook in vijf categorieën te doen, maar dit was niet altijd mogelijk, bijvoorbeeld omdat de factor simpelweg slechts uit twee categorieën bestond. Omdat het multivariate logistisch regressiemodel voor een categorische variabele informatie verschaft of een categorie een significante verhoging of verlaging van de kans op werkhervatting geeft t.o.v. de referentie categorie (d.m.v. de odds ratio, zie par. 3.5) hebben we deze informatie gebruikt om de categorieën van een variabele te ordenen en eventueel samen te voegen. Zie par. 6.1.3 voor een verdere uitleg van de werkwijze.

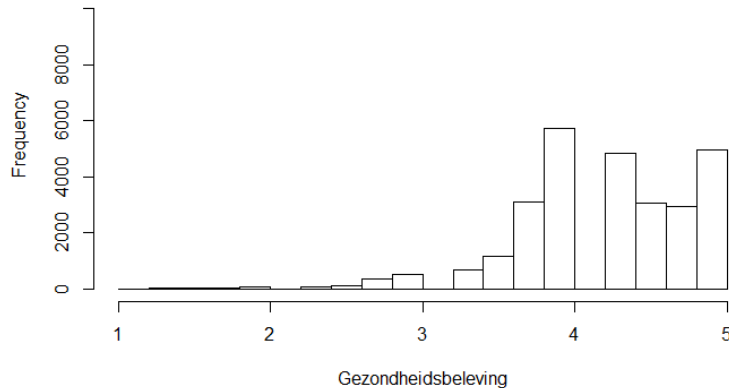
#### 6.1.1.2 Normen voor numerieke factoren

In deze paragraaf leggen we uit hoe we de drie normeringsschalen hebben berekend voor de numerieke factoren. Om de uitleg te vergemakkelijken nemen we als voorbeeld de factor Gezondheidsbeleving (zie tabel 6.1 en figuur 6.1). De score op deze factor is berekend als het gemiddelde op 4 items gemeten op een vijfpuntsschaal.

Tabel 6.1: Frequenties en bijbehorende (cumulatieve) percentages voor de factor Gezondheidsbeleving in de normgroep werkhervatters binnen 12 maanden ( $n = 27.670$ ).

Score	Frequentie	Percentage (%)	Cumulatief	
			Frequentie	Percentage (%)
1	8	0,0	8	0,0
1,25	17	0,1	25	0,1
1,5	26	0,1	51	0,2
1,75	34	0,1	85	0,3
2	50	0,2	135	0,5
2,25	72	0,3	207	0,7
2,5	127	0,5	334	1,2
2,75	342	1,2	676	2,4
3	495	1,8	1.171	4,2
3,25	686	2,5	1.857	6,7
3,5	1.170	4,2	3.027	10,9
3,75	3.120	11,3	6.147	22,2
4	5.729	20,7	11.876	42,9
4,25	4.846	17,5	16.722	60,4
4,5	3.049	11,0	19.771	71,5
4,75	2.930	10,6	22.701	82,0
5	4.969	18,0	27.670	100,0

## Werkhervatting binnen 12 maanden



Figuur 6.1: Histogram van de scores op de factor Gezondheidsbeleving in de normgroep.

### Percentielschaalmethode

Om de waargenomen scores op een factor om te rekenen naar de percentielschaal, ordenen we in de normgroep werkhervatters binnen 12 maanden de waargenomen scores van laag naar hoog en bekijken we de frequentietabel met de cumulatieve percentages (zie tabel 6.1 voor een voorbeeld). Vervolgens bepalen we de percentielen, of percentielrangen, waarbij het gebruikelijk is om een correctie voor continuïteit te passen<sup>24</sup>. De continuïteitscorrectie houdt rekening met het feit dat niet alle scores van een variabele in de normgroep zijn geobserveerd. De percentielrang van een geobserveerde waarde valt daardoor iets lager uit dan het cumulatieve percentage. Het percentiel valt tussen de cumulatieve percentages van de twee geobserveerde scores in zitten. Als de score van een werkzoekende een percentiel van 33 heeft, betekent dat 33% van de werkzoekenden die binnen 12 maanden het werk hebben hervat lager heeft gescoord.

De percentielschaal is vervolgens verdeeld in 5 gelijke delen, of kwintielen, omdat we geïnteresseerd zijn in een indeling in vijven. Het 1<sup>e</sup> kwintiel loopt van percentiel 0 tot 20, het 2<sup>e</sup> kwintiel van 20 tot 40 etc. Tabel 6.2 geeft het resultaat weer voor de factor Gezondheidsbeleving. Als een werkzoekende in het tweede kwintiel scoort, dus in het oranje vakje van tabel 6.2, dan betekent dit dat zijn percentielscore tussen de 20 en 40 ligt. Met ander woorden, 20% van de werkzoekenden die binnen 12 maanden het werk hebben hervat scoorden lager dan hem/haar en 60% van hen scoorden hoger. Een score in het lichtoranje gedeelte interpreteren we daarom als licht belemmerend. Een score op de factor in het lichtgroene gedeelte interpreteren we als licht bevorderend voor de kans op werk hervatting, omdat de percentielen bepaald zijn in de groep werkhervatters.

<sup>24</sup> De percentielen hebben we berekend met de functie `quantile()` in de R software omgeving. We gebruikten voor de continuïteitscorrectie de manier aangeraden door Hyndman en Fan (1996), die voorhanden is in R. We hebben ook andere manieren geprobeerd (bijv. lineaire interpolatie). Deze manieren leverden dezelfde waarden op voor de kwintielen.

Tabel 6.2: Percentielschaal verdeeld in vijf categorieën (i.e. kwintielen) voor de factor Gezondheidsbeleving.

	BELEMMEREND			BEVORDEREND	
Percentiel	0 - 20	21 - 40	41 - 60	61 - 80	81 - 100
Geobserveerde score	≤ 3,75	3,76 - 4,00	4,01 - 4,25	4,26 - 4,75	4,76 – 5,00
Normgroep (rij %)¹	22,2	20,7	17,5	21,6	18,0
Niet-werkhervatters (rij %)²	41,7	19,4	14,0	14,7	10,2

1. Verdeling van de normgroep werkhervatters over de vijf categorieën: bijv. 22,2% scoort in het rode gebied.

2. Verdeling van de groep niet-werkhervatters over de vijf categorieën: bijv. 41,7% scoort in het rode gebied.

In de derde rij van tabel 6.2 geven we aan welk percentage van de normgroep scoort in het betreffende kwintiel (zie ook tabel 6.1). Dus 22,2% van de groep werkhervatters scoort lager of gelijk aan 3,75 op gezondheidsbeleving. Opgeteld zijn de percentages van de rij gelijk aan 100%. In de laatste regel geven we de verdeling over de vijf categorieën van de groep niet-werkhervatters weer. Dus. 41,7% van de niet-werkhervatters scoort in het rode gebied.

### Staninemethode

De stanine staat voor *standard-nine*, deze methode verdeelt de waargenomen scores in 9 standaardcores (1 t/m 9). De term *standard* verwijst naar de standaardnormaal verdeling. Er zijn twee manieren om de stanines te berekenen. De eerste manier maakt gebruik van het gemiddelde en de standaarddeviatie en berekent daarmee standaardcores (i.e. z-scores). De tweede manier maakt gebruik van percentielen en is iets meer geschikt voor scheef verdeelde variabelen. Waarom dit zo is, leggen we uit aan de hand van het centrum van de verdeling. Bij de stanine methode moet het centrum van de verdeling van een factor corresponderen met een standaardcore van 5. Voor scheef verdeelde variabelen weten we dat het 50<sup>e</sup> percentiel, de mediaan, een betere maat is voor het centrum van de verdeling dan het gemiddelde. Daarom kun je bij een scheve verdeling de stanines beter berekenen vanuit percentielen dan vanuit standaardcores. Omdat de waarnemingen op de meeste factoren scheef verdeeld waren (zie bijv. figuur 6.1), gebruikten we deze tweede manier om de standaardcores te bepalen.

Omdat we de percentielen gebruiken, doen we praktisch hetzelfde als bij de percentielschaalmethode. We verdelen de percentielschaal echter op een andere manier. Bij de kwintielen was de verdeling in vijf gelijke delen. De staninemethode verdeelt de percentielschaal in negen delen gebaseerd op de standaardnormaalverdeling, met in het midden de "bulk" van scores en aan de uiteinden slechts een klein percentage. Van deze onderverdeling in negen standaardcores maken we vervolgens een onderverdeling in vijven, die in de praktijk vaak gebruikt wordt:

- > zeer laag (rood), standaardcores 1 en 2, onder het 11<sup>e</sup> percentiel;
- > laag (oranje), standaardcore 3, tussen 11<sup>e</sup>- 23<sup>e</sup> percentiel;
- > gemiddeld (lichtgroen), standaardcores 4 - 6, tussen 23<sup>e</sup>- 77<sup>e</sup> percentiel,
- > hoog (middengroen), standaardcore 7, tussen 77<sup>e</sup>- 89<sup>e</sup> percentiel,
- > zeer hoog (donkergroen), standaardcores 8 en 9, boven 89<sup>e</sup> percentiel.

Het resultaat voor de factor Gezondheidsbeleving ziet er als volgt uit:

Tabel 6.3. Resultaat van de staninemethode voor de factor Gezondheidsbeleving.

	BELEMMEREND			BEVORDEREND	
	1 t/m 2	3	4 t/m 6	7	8 t/m 9
Stanine	1 t/m 2	3	4 t/m 6	7	8 t/m 9
Geobserveerde score	≤ 3,75	3,76 - 4,00	4,01 - 4,75	4,76 – 5,00	-
Normgroep (rij %) <sup>1</sup>	22,2	20,7	39,1	18,0	-
Niet –werkhervatters (rij %) <sup>2</sup>	41,7	19,4	28,7	10,2	-

<sup>1</sup>. Verdeling van de normgroep werkhervatters over de vijf categorieën: bijv. 22,2% scoort in het rode gebied.

<sup>2</sup>. Verdeling van de groep niet-werkhervatters over de vijf categorieën: bijv. 41,7% scoort in het rode gebied.

Als we naar de percentages kijken binnen de normgroep (derde rij tabel 6.3) dan zien we dat de meeste werkzoekenden (39,1%) in het midden scoren (lichtgroen).

### T-schaalmethode

Om op basis van de de T-schaal tot een normverdeling te komen voeren we een lineaire transformatie uit. We zetten eerst de ruwe scores op een factor om in een gestandaardiseerde score, de Z-score waarbij we gebruik maken van het gemiddelde en de standaarddeviatie in de normgroep. We gaan hierbij dus uit van een normaalverdeling. We zullen dit wederom aan de hand van de factor Gezondheidsbeleving uitleggen. Het gemiddelde op deze factor in de normgroep is 4,234 en de standaarddeviatie in de normgroep is 0,580. Als we nu de ruwe score 4 omzetten in een Z-score dan wordt dit:  $(4 - 4,234) / 0,580 = -0,403$ . Vervolgens vermenigvuldigen we deze Z-score met 10 en tellen we er 50 bij op. Dit levert de T-score op. Dus de T-score behorende bij de score 4 wordt:  $10 * -0,403 + 50 = 45,97$ . Doen we dit voor alle scores op een factor, dan zal de T-score een gemiddelde hebben van 50 en een standaarddeviatie (SD) van 10. Om tot een indeling van 5 categorieën te komen gaan we uit van de afkapwaarden in de normgroep behorende bij - 1,5 SD, - 0,5 SD, + 0,5 SD en + 1,5 SD. Dit komt overeen met de volgende grenswaarden op de T-schaal: 35, 45, 55 en 65. Het resultaat voor de factor Gezondheidsbeleving ziet er als volgt uit:

Tabel 6.4: De T-schaal voor de factor Gezondheidsbeleving.

	BELEMMEREND			BEVORDEREND	
	≤ 35,00	35,01–45,00	45,01–55,00	55,01–65,00	≥ 65,01
T-score	≤ 35,00	35,01–45,00	45,01–55,00	55,01–65,00	≥ 65,01
Geobserveerde waarde	≤ 3,36	3,37- 3,94	3,95 - 4,52	4,53 - 5,00	-
Normgroep (rij %) <sup>1</sup>	6,7	15,5	49,2	28,5	-
Niet –werkhervatters (rij %) <sup>2</sup>	20,4	21,3	41,3	17,0	-

<sup>1</sup>. Verdeling van de normgroep werkhervatters over de vijf categorieën: bijv. 6,7% scoort in het rode gebied.

<sup>2</sup>. Verdeling van de groep niet-werkhervatters over de vijf categorieën: bijv. 20,4% scoort in het rode gebied.

Als we nu naar de percentages kijken binnen de groep werkhervatters (derde rij tabel 6.4) dan zien we dat bijna de helft (49,2%) in het midden scoort (lichtgroen). Aan de uiteinden

zijn de percentages lager, met name aan de linkerkant: slechts 7% van de werkhervatters scoort in het rode vakje.

### 6.1.1.3 Normen voor categorische factoren

Om tot een indeling van belemmerend naar bevorderend te komen voor de categorische factoren, hebben we gebruik gemaakt van het multivariate logistische regressie model (gepresenteerd in tabel 3.6, hoofdstuk 3). Om dit logistische regressiemodel te ontwikkelen hebben we de gehele ontwikkeldataset gebruikt ( $n = 26.541$ ). Deze data set gebruiken we nu weer om de normen te ontwikkelen. We gebruiken de status werkhervatting binnen 12 maanden als uitkomstmaat en alle 18 factoren van Werkverkenner 2.0 als voorspellers.

Tabel 6.5: De verdeling van de factor Opleiding in de ontwikkeldataset ( $n = 26.541$ ).

Categorie	Frequentie	%
Laag	5.329	20,1
Midden	13.635	51,4
Hoog	7.577	28,5
Totaal	26.541	100,0

Om te onderzoeken welke categorieën van een factor significant van elkaar verschillen voeren we per factor een aantal analyses uit. Aan de hand van de categorisch factor “Opleiding” leggen we de achtereenvolgende stappen uit van de analyses (zie tabel 6.5 voor de beschrijvende statistieken van deze factor).

**Stap 1.** We voeren in de ontwikkeldataset meerdere malen de logistische regressie analyse uit (met alle 18 factoren), waarbij telkens een andere categorie van de factor die we onderzoeken als referentiecategorie fungeert. We herhalen de analyses totdat iedere categorie een keer de rol van referentiecategorie heeft gehad. Bijvoorbeeld voor de factor Opleiding gebruiken we in de eerste analyse de categorie ‘laag’ als referentie en in de tweede analyse de categorie ‘hoog’. We bekijken vervolgens de odds ratio (OR) en de p-waarde van het effect van iedere categorie t.o.v. de referentiecategorie. De eerste analyse voor de factor Opleiding resulteert in een significant effect voor de categorie ‘midden’ t.o.v. ‘laag’ (OR = 1,19;  $p < 0,001$ ) en een significant effect van de categorie ‘hoog’ t.o.v. ‘laag’ (OR = 1,10;  $p = 0,043$ ; zie ook tabel 3.6). De tweede analyse resulteert in een significant effect van de categorie “midden” t.o.v. “hoog” (OR = 1,08;  $p = 0,03$ ).

**Stap 2.** Als de odds ratio van een categorie van een factor t.o.v. de referentiecategorie niet significant is, betekent dit dat het effect van de twee categorieën op de kans op werkhervatting vrijwel gelijk is. Vanwege de grote steekproef in onze data set betekende dit feitelijk dat de odds ratio meestal rond de 1 lag. Als dit het geval was, voegden we de categorieën samen. Voor de factor Opleiding voegen we geen categorieën samen omdat alle odds ratio's significant zijn.

**Stap 3.** Vervolgens voeren we nogmaals een logistische regressie uit met de resulterende indeling uit de vorige stap. Op basis van de grootte van de odds ratio's ordenen we de categorieën van belemmerend naar bevorderend. Voor de factor Opleiding resulteert deze ordening in: laag, hoog, midden. Zoals we al eerder zagen in hoofdstuk 3 interpreteren we de odds ratio als de toename in kans op werkhervatting. We stelden bijvoorbeeld dat werkzoekenden met een opleidingsniveau in de midden categorie een 1,19 keer zo grote kans op werkhervatting hebben als werkzoekenden met een laag opleidingsniveau. De hoogte van de odds ratio geeft dus een indicatie hoe bevorderend de categorie is voor de

kans op werkhervatting t.o.v. een andere categorie. In onze uiteindelijke indeling in kleuren geven de opeenvolgende kleuren een indicatie hoe bevorderend een categorie is voor de kans op werkhervatting t.o.v. een andere categorie. Om de kleurenindeling te bepalen hanteren we de volgende regels:

- > Als de odds ratio van 2 opeenvolgende categorieën lager is dan 1,5, dan verschillen de categorieën slechts 1 kleur (dus bijvoorbeeld van oranje naar lichtgroen).
- > Als we minder dan vijf categorieën hebben (na stap 3), dan geven we de niet gebruikte kleuren lichtgrijs weer.
- > We nemen de categorie die het meest belemmerend is voor de kans op werkhervatting als uiteindelijke referentiecategorie (voor de odds ratio's die we presenteren in de tabel).
- > Bij vier of vijf categorieën valt de referentiecategorie in het rode vak. Bij minder dan vier categorieën valt de referentiecategorie in het rode, oranje, of lichtgroene vak (dit is bepaald in overleg met UWV).

Het resultaat voor de factor Opleiding is weergegeven in tabel 6.6. Een voordeel van de methode op basis van de odds ratio's is dat categorieën die van kleur verschillen ook significant verschillen t.o.v. elkaar in het effect op werkhervatting. Dit kunnen we niet stellen voor de numerieke factoren. Voor de numerieke factoren geldt echter al bij voorbaat dat alle odds ratio's significant zijn in het logistisch regressiemodel (zie tabel 3.6). Afhankelijk van de waarde van de odds ratio (hoger of lager dan 1) weten we dat een hogere score op de numerieke factor meer bevorderend (of belemmerend) is voor de kans op werkhervatting.

Tabel 6.6: De indeling van belemmerend naar bevorderend op basis van de odds ratio's in het multivariate logistisch regressiemodel voor de factor Opleiding

	BELEMMEREND			BEVORDEREND	
Geobserveerde waarde		laag	hoog	midden	
Odds ratio		1,00	1,10	1,19	
Normgroep (rij %) <sup>1</sup>		17,9	28,6	53,5	
Niet-werkhervatters (rij %) <sup>2</sup>		22,3	28,6	49,0	

1. Verdeling van de normgroep werkhervatters over de drie categorieën.

2. Verdeling van de groep niet-werkhervatters over de drie categorieën.

### 6.1.2 Resultaten van de normeringsschalen

De resultaten van de drie normeringsschalen voor numerieke factoren hebben we in samenspraak met UWV vergeleken op bruikbaarheid in de praktijk. Belangrijkste criteria hierbij waren: 1) de indeling van de scores in vijf categorieën geeft goed aan wanneer de factor belemmerend wordt; 2) de indeling is goed uit te leggen aan de adviseur, met name bij welke score er sprake van een belemmering is. Wat betreft het eerste criterium: de taak van de adviseur is om extra aandacht te geven aan de factoren waarop een werkzoekende belemmerend scoort. Het is daarbij niet wenselijk dat een score op een factor voor een relatief grote groep werkhervatters in het rode gebied valt. Dan heeft dit minder betekenis. Op basis van deze overweging hebben de staninemethode en de T-schaal de voorkeur boven de kwintielen. Bij deze laatste methode is het percentage werkhervatters dat in het rode gebied valt immers hoger dan bij de andere twee. Wat betreft het tweede criterium: de indelingen op basis van de staninemethode en op basis van de T-schaal waren voor de meeste factoren hetzelfde en ook goed uit te leggen aan de adviseurs. Echter voor

bijvoorbeeld de factor Gezondheidsbeleving had de T-schaal de voorkeur boven de staninemethode. Het is lastig om te verantwoorden dat een score van 4 op deze factor, een goede gezondheidsbeleving, al in het oranje, licht belemmerende, gebied valt, zoals bij de staninemethode het geval was (zie tabel 6.3). Op basis van deze praktische overwegingen is de T-schaal als meest bruikbare schaal naar voren gekomen. We beperken ons daarom tot het beschrijven van de resultaten van de T-schaal. In bijlage H worden de resultaten van de andere twee methodes getoond.

Hieronder zijn de resultaten van 16 van de 18 factoren van Werkverkenner 2.0 weergegeven. Twee van de factoren ontbreken op verzoek van UWV: Soort bemiddelingsberoep en Bedrijfstak. UWV maakt actuele arbeidsmarktinformatie over kansrijke beroepen en bedrijfstakken die goed bruikbaar is voor de adviseur werk bij het stellen van een diagnose. Omdat deze informatie op reguliere basis geactualiseerd wordt door UWV, zal de adviseur voor de diagnose van de factoren beroep en bedrijfstak hiervan gebruik blijven maken. Voorbeelden hiervan staan op: [https://www.werk.nl/werk\\_nl/\\_arbeidsmarktinformatie/sector-beroep](https://www.werk.nl/werk_nl/_arbeidsmarktinformatie/sector-beroep)

We hebben de 16 factoren geordend naar de hoogte van de bijdrage aan de voorspelling van werkhervatting (dus op basis van de volgorde van tabel 3.6). Bij deze manier van presenteren komen numerieke factoren en categorische factoren door elkaar te staan. Naast de verdeling van de groep werkhervatters over de vijf kleuren geven we ook per factor de verdeling van de groep niet-werkhervatters en de verdeling van de totale groep. Dit is om snel een inzicht te geven in bijvoorbeeld hoe groot het percentage werkzoekenden is die belemmerend scoort binnen de verschillende groepen.

De eerstgenoemde factor, Leeftijd, heeft het sterkste effect op werkhervatting. Merk op dat Leeftijd een kansverkleinende factor is, daarom zijn de kritieke waardes op de T-schaal omgedraaid,  $\geq 65,01$  is rood. We geven alleen van de categorische factoren de odds ratio weer. Voor de odds ratios van de numerieke factoren verwijzen we naar tabel 3.6.

Tabel 6.7: Leeftijd (één item; numeriek)

Leeftijd	BELEMMEREND			BEVORDEREND	
	$\geq 65,01$	55,01–65,00	45,01–55,00	35,01–45,00	$\leq 35,00$
T-score	$\geq 65,01$	55,01–65,00	45,01–55,00	35,01–45,00	$\leq 35,00$
Geobserveerde score	$\geq 55$	44 - 54	33 - 43	22 - 32	$\leq 21,00$
Normgroep (rij %) <sup>1</sup>	8,6	25,7	28,9	34,7	2,1
Niet –werkhervatters (rij %) <sup>2</sup>	28,2	32,9	25,6	13,0	0,2

<sup>1</sup>. Verdeling van de normgroep werkhervatters over de vijf categorieën: bijv. 8,6% scoort in het rode gebied.

<sup>2</sup>. Verdeling van de groep niet-werkhervatters over de vijf categorieën: bijv. 28,2% scoort in het rode gebied.

Tabel 6.8: Aantal jaren in laatste functie (één item; categorisch)

	BELEMMEREND			BEVORDEREND	
Geobserveerde score	10 t/m 50	5,1 t/m 10	1,1 t/m 5,0	0,6 t/m 1,0	0 t/m 0,5
Odds ratio	1,00	1,20	1,67	1,93	2,17
Normgroep (rij %) <sup>1</sup>	12,7	13,0	43,1	14,9	16,3
Niet –werkhervatters (rij %) <sup>2</sup>	31,1	18,5	33,2	8,7	8,4

<sup>1</sup>. Verdeling van de normgroep werkhervatters over de vijf categorieën.

<sup>2</sup>. Verdeling van de groep niet-werkhervatters over de vijf categorieën.

Tabel 6.9: Visie op terugkeer naar werk (gemiddelde van drie items op vijfpuntschaal; numeriek)

	BELEMMEREND			BEVORDEREND	
T-score	≤ 35,00	35,01–45,00	45,01–55,00	55,01–65,00	≥ 65,01
Geobserveerde score	≤ 2,57	2,58 - 3,30	3,31 - 4,04	4,05 - 4,77	4,78 - 5,00
Normgroep (rij %) <sup>1</sup>	5,4	18,9	52,6	15,6	7,4
Niet –werkhervatters (rij %) <sup>2</sup>	15,5	29,3	45,2	7,8	2,1

<sup>1</sup>. Verdeling van de normgroep werkhervatters over de vijf categorieën.

<sup>2</sup>. Verdeling van de groep niet-werkhervatters over de vijf categorieën.

Tabel 6.10: Samenstelling huishouden (samengesteld van twee items; categorisch); getrouwd bevat ook de categorie “woont samen”.

	BELEMMEREND			BEVORDEREND	
Geobserveerde categorie		Woont alleen; Vrouw, getrouwd, jongste kind thuis 0-6 jaar; Alleen- staande ouder; Anders.	Getrouwd, geen thuiswonende kinderen; Man getrouwd, jongste kind thuis 0-6 jaar.	Getrouwd, jongste kind thuis 7 jaar en ouder.	
Odds ratio		1,00	1,10	1,19	
Normgroep (rij %) <sup>1</sup>		46,4	31,9	21,7	
Niet –werkhervatters (rij %) <sup>2</sup>		41,2	34,7	24,1	

<sup>1</sup>. Verdeling van de normgroep werkhervatters over de drie categorieën.

<sup>2</sup>. Verdeling van de groep niet-werkhervatters over de drie categorieën.



Tabel 6.11: Maximale duur WW (één item; categorisch)

	BELEMMEREND		BEVORDEREND	
		> 12 mnd		< = 12 mnd
Geobserveerde categorie				
Odds ratio		1,00		1,69
Normgroep (rij %) <sup>1</sup>		65,0		35,0
Niet –werkhervatters (rij %) <sup>2</sup>		87,4		12,6

1. Verdeling van de normgroep werkhervatters over de twee categorieën.

2. Verdeling van de groep niet-werkhervatters over de twee categorieën.

Tabel 6.12: Werkzoekgedrag m.b.t. direct contact werkgevers (gemiddelde van drie items op vierpuntsschaal; numeriek)

	BELEMMEREND			BEVORDEREND	
	≤ 35,00	35,01–45,00	45,01–55,00	55,01–65,00	≥ 65,01
T-score					
Geobserveerde score	≤ 1,11	1,12 – 1,73	1,74 – 2,35	2,36 – 2,97	2,98 – 4,00
Normgroep (rij %) <sup>1</sup>	8,0	29,2	40,9	10,6	11,4
Niet –werkhervatters (rij %) <sup>2</sup>	17,8	38,0	31,6	6,7	6,0

1. Verdeling van de normgroep werkhervatters over de vijf categorieën.

2. Verdeling van de groep niet-werkhervatters over de vijf categorieën.

Tabel 6.13: Gezondheidsbeleving (gemiddelde van vier items op een vijfpuntsschaal; numeriek)

	BELEMMEREND			BEVORDEREND	
	≤ 35,00	35,01–45,00	45,01–55,00	55,01–65,00	≥ 65,01
T-score					
Geobserveerde waarde	≤ 3,36	3,37- 3,94	3,95 - 4,52	4,53 - 5,00	-
Normgroep (rij %) <sup>1</sup>	6,7	15,5	49,2	28,5	-
Niet –werkhervatters (rij %) <sup>2</sup>	20,4	21,3	41,3	17,0	-

1. Verdeling van de normgroep werkhervatters over de vijf categorieën.

2. Verdeling van de groep niet-werkhervatters over de vijf categorieën.

Tabel 6.14: Balans voor- en nadelen niet-werken (één item; categorisch)

Geobserveerde categorie	BELEMMEREND		BEVORDEREND		
		evenveel voordelen als nadelen of meer voordelen dan nadelen	meer nadelen dan voordelen		
Odds ratio		1,00	1,24		
Normgroep (rij %) <sup>1</sup>		18,9	81,1		
Niet –werkhervatters (rij %) <sup>2</sup>		31,4	68,6		

1. Verdeling van de normgroep werkhervatters over de twee categorieën.

2. Verdeling van de groep niet-werkhervatters over de twee categorieën.

Tabel 6.15: WIA 35-min-indicatie (één item; categorisch)

Geobserveerde categorie	BELEMMEREND		BEVORDEREND		
	WIA 35-min		Geen WIA 35-min		
Odds ratio	1,00		2,02		
Normgroep (rij %) <sup>1</sup>	0,8		99,2		
Niet –werkhervatters (rij %) <sup>2</sup>	3,4		96,6		

1. Verdeling van de normgroep werkhervatters over de twee categorieën.

2. Verdeling van de groep niet-werkhervatters over de twee categorieën.

Tabel 6.16: Problemen Nederlandse taal (één item; categorisch)

Maximale duur WW	BELEMMEREND		BEVORDEREND		
		Kan geen NL / veel of beetje problemen	Geen problemen		
Odds ratio		1,00	1,38		
Normgroep (rij %) <sup>1</sup>		5,6	94,4		
Niet –werkhervatters (rij %) <sup>2</sup>		5,8	94,2		

1. Verdeling van de normgroep werkhervatters over de twee categorieën.

2. Verdeling van de groep niet-werkhervatters over de twee categorieën.

Tabel 6.17: Acceptatiebereidheid m.b.t. voltijd baan (één item op vijf puntsschaal; numeriek)

	BELEMMEREND			BEVORDEREND	
T-score	≤ 35,00	35,01–45,00	45,01–55,00	55,01–65,00	≥ 65,01
Geobserveerde waarde	1	2 - 3	4	5	-
Normgroep (rij %) <sup>1</sup>	12,2	17,0	17,7	53,1	-
Niet –werkhervatters (rij %) <sup>2</sup>	23,9	25,1	17,4	33,6	-

<sup>1</sup>. Verdeling van de normgroep werkhervatters over de vijf categorieën.

<sup>2</sup>. Verdeling van de groep niet-werkhervatters over de vijf categorieën.

Tabel 6.18: Algemeen werkvermogen (één item op tien puntsschaal; numeriek)

	BELEMMEREND			BEVORDEREND	
T-score	≤ 35,00	35,01–45,00	45,01–55,00	55,01–65,00	≥ 65,01
Geobserveerde waarde	≤ 6	7	8 – 9	10	-
Normgroep (rij %) <sup>1</sup>	7,3	13,0	57,7	22,1	-
Niet –werkhervatters (rij %) <sup>2</sup>	20,9	18,1	48,1	13,0	-

<sup>1</sup>. Verdeling van de normgroep werkhervatters over de vijf categorieën: bijv. 7,3% scoort in het rode gebied.

<sup>2</sup>. Verdeling van de groep niet-werkhervatters over de vijf categorieën: bijv. 20,9% scoort in het rode gebied.

Tabel 6.19: Aantal uren in staat te werken (één item, 0 – 50 uur per week; numeriek)

	BELEMMEREND			BEVORDEREND	
T-score	≤ 35,00	35,01–45,00	45,01–55,00	55,01–65,00	≥ 65,01
Geobserveerde waarde	≤ 22	23 - 32	33 - 42	43 - 50	-
Normgroep (rij %) <sup>1</sup>	7,3	21,6	44,5	26,5	-
Niet –werkhervatters (rij %) <sup>2</sup>	19,0	31,2	37,1	12,7	-

<sup>1</sup>. Verdeling van de normgroep werkhervatters over de vijf categorieën: bijv. 7,3% scoort in het rode gebied.

<sup>2</sup>. Verdeling van de groep niet-werkhervatters over de vijf categorieën: bijv. 19,0% scoort in het rode gebied.

Tabel 6.20: Opleidingsniveau (één item; categorisch)

	BELEMMEREND			BEVORDEREND	
Geobserveerde categorie		laag	hoog	midden	
Odds ratio		1,00	1,10	1,19	
Normgroep (rij %) <sup>1</sup>		17,9	28,6	53,5	
Niet –werkhervatters (rij %) <sup>2</sup>		22,3	28,6	49,0	

<sup>1</sup>. Verdeling van de normgroep werkhervatters over de drie categorieën.

<sup>2</sup>. Verdeling van de groep niet-werkhervatters over de drie categorieën.

Tabel 6.21: Acceptatiebereidheid m.b.t. afwijkende tijden (gemiddelde van twee items op een vijfpuntsschaal: zeker niet tot zeker wel)

	BELEMMEREND			BEVORDEREND	
T-score	≤ 35,00	35,01–45,00	45,01–55,00	55,01–65,00	≥ 65,01
Geobserveerde score	≤ 1	1,1 – 2,2	2,3 - 3,5	3,6 - 4,7	4,8 – 5,0
Normgroep (rij %) <sup>1</sup>	12,3	22,1	36,8	17,1	11,7
Niet –werkhervatters (rij %) <sup>2</sup>	22,2	25,8	33,6	11,6	6,9

1. Verdeling van de normgroep werkhervatters over de vijf categorieën.

2. Verdeling van de groep niet-werkhervatters over de vijf categorieën.

Tabel 6.22: Inkomsten uit werk naast WW (één item; categorisch)

	BELEMMEREND		BEVORDEREND	
Geobserveerde categorie			Nee	Ja
Odds ratio			1,00	1,17
Normgroep (rij %) <sup>1</sup>			84,6	15,4
Niet –werkhervatters (rij %) <sup>2</sup>			86,7	13,3

1. Verdeling van de normgroep werkhervatters over de twee categorieën.

2. Verdeling van de groep niet-werkhervatters over de twee categorieën.

### 6.1.3 Impact van beïnvloeding van een factor op de kans op werkhervatting

Wanneer we voor alle variabelen de regressiecoëfficiënten van het uiteindelijke model invullen in een predictieregel, dan kunnen we voor iedere werkzoekende de kans uitrekenen dat deze persoon het werk heeft hervat binnen 12 maanden. Voor een werkzoekende uit de onderzoeksgroep, hier genaamd 'Pierre', hebben we de predictieregel ingevuld met zijn persoonlijke scores uit de Werkverkenner. Deze Pierre is ook als casus 1 gebruikt in het gebruikersonderzoek bij de adviseurs werk (zie de rapportage gebruikersonderzoek, hoofdstuk 3). Pierre is een getrouwde man van 59 jaar en heeft geen thuiswonende kinderen. Wanneer we voor Pierre de scores op alle factoren in de predictieregel invullen, dan is zijn voorspelde kans op werkhervatting op basis van deze predictieregel 12%. Dit betekent dat hij een lage kans heeft om het werk binnen 12 maanden te hervatten. Als we kijken naar de individuele scores op de factoren van Pierre, dan geeft dat inzicht in waar hij zit ten opzichte van de populatie WW-kanten die het werk hervatten en waar mogelijk op kan worden ingesprongen om zijn kans op werkhervatting binnen 12 maanden te vergroten. Tabel 6.23 geeft een weergave van de scores op alle factoren van Pierre die mogelijk getoond kan worden als uitslag aan de adviseur werk.

We gaan hier nader in op twee belangrijke kansvergrotende factoren uit het uiteindelijke model, te weten Visie op terugkeer naar werk en Gezondheidsbeleving. Pierre heeft een score van 2,0 op Visie op terugkeer naar werk. Zijn score valt dus in het rode gebied van deze factor (tabel 6.9). We kunnen ook bepalen wat een verhoging van de score van Pierre betekent voor zijn kans op werkhervatting binnen 12 maanden. Zou hij bijvoorbeeld (of dit realistisch is, laten we hier buiten beschouwing) drie punten omhoog gaan in zijn score op Visie terugkeer naar werk (van 2,0 naar 5,0), dan komt hij terecht in het donkergroene vakje

en vergroot hij zijn kans op werkhervatting binnen 12 maanden van 12% naar 27%. Dit is dus een absolute toename van 15%.

Hetzelfde kunnen we doen voor de factor Gezondheidsbeleving. Pierre heeft een score van 3,0 op Gezondheidsbeleving, matig, en heeft hiermee een score in het rode vak (tabel 6.13). Stel dat hij zijn gezondheidsbeleving kan vergroten naar een score van 5 (dat betekent naar zeer goed), dan vergroot hij zijn kans op werkhervatting binnen 12 maanden van 12% naar 18%, een toename van 6%.

Tabel 6.23: Uitslag van een voorbeeld casus (Pierre) op de factoren van Werkverkenner 2.0 (m.u.v. Soort bemiddelingsberoep en Bedrijfstak). De scores van Pierre zijn door middel van een X weergegeven.

	BELEMMEREND		BEVORDEREND		
Leeftijd	X				
Aantal jaren in laatste functie			X		
Visie op terugkeer naar werk	X				
Samenstelling huishouden			X		
Maximale duur WW		X			
Werkzoekgedrag: Direct contact	X				
Gezondheidsbeleving	X				
Balans voor- en nadelen		X			
WIA 35-min			X		
Problemen Nederlands taal			X		
Acceptatie m.b.t. voltijdbaan	X				
Algemeen werkvermogen	X				
Aantal uren in staat te	X				
Opleidingsniveau			X		
Acceptatiebereidheid afw.	X				
Inkomsten uit werk			X		

Noot. Kleurvakken die niet voor kunnen komen zijn in grijs getoond.

De toename in kans op werkhervatting is niet voor iedere casus hetzelfde. Dit komt door het niet-lineaire verband tussen de voorspelde log(odds) door het logistisch regressiemodel en de voorspelde kans. Ons voorbeeld Pierre is een casus met een lage kans op werkhervatting. Stel dat we uitgaan van een gemiddelde casus, d.w.z. een werkzoekende die op alle numerieke factoren van Werkverkenner 2.0 gemiddeld scoort en op alle categorische factoren rond de modus. Om de invloed van Visie op terugkeer naar werk te bepalen, geven we aan deze casus een rode score, bijvoorbeeld 2,33. Zijn kans op werkhervatting is in dit geval 41%. Verhogen we vervolgens deze score van 2,33 naar 5, dus naar donkergroen, dan wordt zijn kans op werkhervatting 63%. Dat is een toename van 22%. Dit is meer dan de absolute toename van 15% die we bij Pierre zagen. Vervolgens zetten we de score op Visie terugkeer naar werk weer op gemiddeld en zetten we de factor Gezondheidsbeleving op rood, bijvoorbeeld 3. Zijn kans op werkhervatting is dan 49%. Als we vervolgens de score op Gezondheidsbeleving verhogen naar donkergroen, dan wordt zijn kans op werkhervatting 71%. De toename in zijn kans op werkhervatting is dan 23%, bij Pierre was de toename slechts 6%.

We concluderen dat het inderdaad mogelijk is om de toename van de kans zichtbaar te maken voor individuele werkzoekenden. Het effect van de toename van een score op een factor, bijvoorbeeld van rood naar donkergroen, op de kans op werkherhvatting hangt af van het gehele profiel van de werkzoekende (d.w.z. zijn/haar scores op alle andere factoren van Werkverkenner 2.0). Als UWV daar behoefte aan heeft, kan de toename van de kans worden ingebouwd in de terugkoppeling aan de adviseur werk.

## 6.2 Gebruik van de Werkverkenner in de praktijk

Naast een goede normering (zie par. 6.1) is ook de gebruiksvriendelijkheid van de toepassing van de uitkomsten van de Werkverkenner van belang om een goede snelle diagnose te kunnen stellen. Daarom is er onderzoek verricht onder gebruikers. Het gebruikersonderzoek richt zich specifiek op de twee typen gebruikers van de Werkverkenner: de WW-klant en de adviseurs werk. De resultaten van het gebruikersonderzoek zijn uitgebreid beschreven in een aparte rapportage met gerichte aanbevelingen met betrekking tot de software en technische werking, inhoud en feedback (Kranenburg, Dusseldorp & Sonke, 2017). Deze aparte rapportage is bestemd voor de UWV specialisten die de Werkverkenner gaan implementeren, waarbij ze rekening moeten houden met de gebruiksvriendelijkheid van de Werkverkenner voor zowel werkzoekenden als adviseurs binnen de context van de gehele dienstverlening. In deze paragraaf gaan we kort in op de belangrijkste bevindingen van deze gebruikers en geven aanbevelingen met betrekking tot de inhoud en feedback van de Werkverkenner zelf.

Het *doel* van het gebruikersonderzoek is om de aspecten van het gebruik van de Werkverkenner in kaart te brengen. We richten ons op het beantwoorden van de volgende vragen:

- > Zijn de instructies duidelijk voor de WW-klienten?
- > Zijn de vragen begrijpelijk voor de WW-klienten?
- > Is het invullen van de Werkverkenner gebruiksvriendelijk?
- > Is de rapportage na invulling van de vragenlijst voor zowel de WW-klant als de adviseurs werk klantvriendelijk?

### *User-walkthrough met WW-klienten*

Om het gehele proces van het invullen van de Werkverkenner in kaart te brengen, hebben we in dit gebruikersonderzoek gekozen voor een kwalitatief onderzoek in de vorm van een *user-walkthrough* met een beperkt aantal WW-klienten. Deze vorm biedt de mogelijkheid om de WW-klant te observeren tijdens het invullen van de vragenlijst, open vragen te stellen en om door te vragen, om zo een beter beeld te krijgen wat de klant bedoelt.

### *Focusgroep met adviseurs werk*

Om meningen en visies van de adviseurs werk over de rapportage van de uitslag van een WW-klant op de Werkverkenner in kaart te brengen is gekozen voor een *focusgroep*. Dit is een vorm van kwalitatief onderzoek waarbij een groep mensen gevraagd wordt naar hun meningen. Deze vorm heeft als voordeel dat er discussie mogelijk is. Hierdoor krijgt de onderzoeker snel een beeld over welke onderwerpen de groep het eens of juist oneens is en welke onderwerpen belangrijk zijn. Tijdens de focusgroep werd de huidige rapportage naast verschillende andere mogelijke toekomstige rapportages, scenario's, gelegd.

## 6.2.1 Gebruikersonderzoek onder WW-klienten

### 6.2.1.1 Onderzoeksopzet

Het gebruikersonderzoek van WW-klienten is uitgevoerd onder acht werkzoekenden. De deelnemers aan de user-walkthrough zijn via UWV geworven onder WW-klienten die recent een verzoek hadden gekregen om de Werkverkenner in te vullen in de regio Den Haag. Het verzoek tot vrijwillige deelname werd gedaan via een digitaal bericht in de Werkmap, de persoonlijke digitale omgeving bij UWV van de werkzoekende, met daarin een beloofde vergoeding van 30 euro. Dit digitaal bericht werd geplaatst op twee momenten: de eerste keer bij 100 WW-klienten en vervolgens twee weken later bij 390 WW-klienten. Wanneer men zich wilde aanmelden voor de user-walkthrough, dan kon men contact opnemen met TNO. Uiteindelijk meldden tien WW-klienten zich aan, waarvan 2 op de reservelijst werden gezet. Toen het vooraf bepaalde quotum van acht WW-klienten was behaald, is gestopt met werven.

De leeftijd van de deelnemers aan de user-walkthrough varieert van 28 tot 60 jaar. De gemiddelde leeftijd was 40 jaar, waarvan vier personen jonger dan 35 jaar en vier personen ouder dan 41 jaar waren. Er deden twee vrouwen en zes mannen mee. De nationaliteit van alle acht deelnemers was Nederlands. Het opleidingsniveau varieerde van havo (één persoon), mbo (twee personen), hbo (drie personen) tot wetenschappelijk onderwijs (één persoon). Alle deelnemers hebben een computer thuis en een mobiele telefoon en ieder maakt hier dagelijks gebruik van.

Het onderzoek bestond uit een individuele sessie met iedere deelnemer onder begeleiding van een proefleider van TNO. Een sessie duurde maximaal twee uur. De sessies vonden plaats in een vergaderruimte bij TNO, locatie Den Haag, op 10 en 15 maart 2016.

De opbouw van de sessie was als volgt:

- > *Interview en vragenlijst.* Hierin werd gevraagd naar demografische gegevens, ervaring met ICT en ervaringen met UWV en de Werkverkenner.
- > *User-walkthrough.* Tijdens de user-walkthrough vulde de deelnemer de Werkverkenner in. De handelingen die daarvoor nodig waren, werden zoveel mogelijk zelfstandig door de deelnemer uitgevoerd. Het gaat hierbij om het openen en invullen van de vragenlijst, het afhandelen van foutmeldingen en het gebruik van hulpfuncties. De proefleider zat naast de deelnemer en hielp de hem/haar indien dit nodig was. Tijdens het invullen van de Werkverkenner observeerde de proefleider de handelingen en noteerde de bevindingen op een observatieformulier.
- > *Interview.* Na het invullen van de Werkverkenner vroeg de proefleider in een interview naar de ervaringen van de deelnemer. Dit gebeurde aan de hand van criteria, waarbij onder andere de aspecten begrijpelijkheid van de vragen, bruikbaarheid van de Werkverkenner en de terugkoppeling aan bod kwamen.


Voor de user-walkthrough werd de methode 'cognitive walkthrough' gebruikt. Dit is een methode uit de ergonomie om de gebruiksvriendelijkheid te evalueren. Hiervoor voeren één of meerdere gebruikers een serie taken uit en beantwoorden een aantal vragen. Deze methode wordt gebruikt om met aandacht voor het geheel, problemen in de gebruiksvriendelijkheid van een interactief systeem op te sporen. De focus ligt hierbij op het gemak waarmee nieuwe en ervaren gebruikers taken met het systeem uitvoeren en het systeem leren kennen. Tijdens een sessie voerde een deelnemer een aantal relevante taken uit (tabel 6.24) met als doel gebruikerservaringen en mogelijke knelpunten te inventariseren en waar mogelijk verbeterpunten te formuleren.

Tabel 6.24: Handelingen uitgevoerd door de deelnemer tijdens de user-walkthrough.

Opdrachten uitgevoerd door de respondent tijdens de user-walkthrough	
>	Log in op uw werkmap
>	Zoek taken / nieuwe items in uw werkmap
>	Vul de vragenlijst in
>	Ga één vraag terug
>	Aangeven hoe men de volgende acties zou uitvoeren: bewaren en stoppen van de vragenlijst, uitloggen, verlengen van uw sessie, gebruiken van de hulpfunctie

Om knelpunten en aanbevelingen te categoriseren maken we tijdens de sessie gebruik van COTAN<sup>25</sup> criteria (Evers et al., 2010) en ergonomische ontwerpprincipes (Neerincx & Lindenberg 1999, Neerincx et al. 1999, Neerincx et al. 2001). Om het proces te begrijpen dat iemand doorloopt tijdens het beantwoorden van vragen in een vragenlijst en te begrijpen op welke punten het fout gaat en waarom werd het model van Tourangeau (Tourangeau et al. 2000) gebruikt.

Tabel 6.25: Model van Tourangeau.

There are four phases in answering a question		
Phase	A good question...	
1. Read and understand (comprehend the question by identifying cues)	Is legible and makes sense	
2. Find an answer (retrieve relevant information from memory)	Asks for answers that we know	
3. Judge the answer (Integrate available information into judgement)	Asks for answers we're happy to reveal	
4. Place the answer (map judgement on response and give an answer)	Offers appropriate spaces for the answer	

Volgens dit model, moet het menselijk brein vier fasen doorlopen om een passend antwoord te geven op een vraag (zie tabel 6.25). Met behulp van dit model kan geanticipeerd worden op het mentale proces van de respondent bij de beantwoording van de vragen.

Om te testen hoe duidelijk de Werkverkenner is voor de deelnemers, werd de Three Step Test Interview methode van Hak en van der Veer (Hak et al., 2004) gebruikt. Dit is een vorm van cognitief interviewen, die bestaat uit verschillende stappen. In stap 1 denkt iemand hardop terwijl hij de vragenlijst invult. Vervolgens wordt in stap 2 informatie aangevuld, die tijdens stap 1 is gemist, bijvoorbeeld doordat iemand even vergat hardop te denken. In stap 3 wordt de persoon geïnterviewd over wat er tijdens stap 1 gebeurde. Gevraagd wordt naar wat daar mogelijk aan ten grondslag lag, hoe de persoon het zelf heeft ervaren en of de persoon mogelijk suggesties heeft voor verbetering.

### Bevindingen

Het starten en invullen van Werkverkenner 1.0+ gaat vlot en duurt ongeveer 7-15 minuten. Dit vinden de deelnemers kort. Hierbij dient te worden opgemerkt dat Werkverkenner 1.0+ met 50 vragen de meest uitgebreide versie van de Werkverkenner is. Werkverkenner 2.0,

<sup>25</sup> Commissie Testaangelegenheden Nederland. Voor de criteria zie Evers et al 2010)



waartoe dit onderzoek is uitgevoerd, zal korter zijn en uit 23 vragen bestaan<sup>26</sup>. Het is dus goed nieuws dat deze lange versie als kort wordt ervaren.

Ook de taken, zoals het bedienen van en het navigeren in de vragenlijst en het zoeken naar hulp, verlopen over het algemeen zonder problemen en worden als prima, simpel en helder benoemd. Gemiddeld beoordeelden de deelnemers de Werkverkenner met het rapportcijfer 7,6. Tabel 6.26 geeft aan hoe de deelnemers de Werkverkenner waarden op basis de diverse kwaliteitscriteria die gesteld worden aan test en vragenlijsten.

Tabel 6.26: Waardering van WW-klienten op verschillende criteria van de Werkverkenner:

Criterion	goed	voldoende	onvoldoende
1. Zijn de items vrij van racistische, ethnocentrische, seksistische en voor bepaalde bevolkingsgroepen kwetsende inhoud?	8 respondenten		
2. Is de software zodanig ontworpen dat fouten door onjuist gebruik kunnen worden vermeden?	7 respondenten	1 respondent	
3. Is de instructie voor de deelnemer volledig en duidelijk?	5 respondenten	3 respondenten	
4. Zijn de items correct geformuleerd?	5 respondenten	3 respondenten	
5. Hoe is de kwaliteit van de vormgeving van de gebruikersinterface?	5 respondenten	3 respondenten	
6. Is de test voldoende veilig	4 respondenten	3 respondenten	1 respondent

De deelnemers vinden de vragenlijst niet kwetsend (criterium 1). Ze vinden de vragenlijst duidelijk en als er zich een probleem voordoet, dan merkt men op dat dit over het algemeen eenvoudig op te lossen is (criterium 2). Via de website van UWV wordt op verschillende manieren ondersteuning en hulp geboden. Er staan veel gestelde vragen en het is mogelijk telefonisch contact op te nemen met een helpdesk. De Hulpknop (?-icoon) in de werkmap wordt door de deelnemers niet genoemd als een optie bij vastlopen tijdens het invullen van de vragen. Andere mogelijkheden tot hulp krijgen, bellen, zoeken via zoekbalk, gerichte vraag stellen aan de adviseur werk via de berichten inbox in de Werkmap, handleiding raadplegen enzovoorts, worden wel genoemd. Wanneer de proefleider de deelnemers wijst op het icoon begrijpt men waarvoor dit bedoeld is.

De instructie wordt over het algemeen ervaren als volledig en duidelijk (criterium 3). Het taalgebruik vindt men eenvoudig. Verder worden de leesbaarheid en de lettergrootte van de teksten als goed ervaren. Een deelnemer benoemt dat het erg fijn is dat hij/zij een ontvangstbevestiging ontvangt. Ook de meeste vragen en de antwoordcategorieën uit de Werkverkenner zijn duidelijk (criterium 4), volgens de deelnemers. Diverse deelnemers geven aan dat zij behoefte hebben om extra verduidelijking te geven via een open veld. Bij een paar vragen vallen de meeste deelnemers over de formuleringen. De meeste van deze vragen komen niet meer terug in Werkverkenner 2.0, omdat de bijbehorende factoren niet voorspellend bleken te zijn. Twee voorbeelden van een dergelijke vraag, zijn de vragen die beginnen met “Geef aan of de situatie voor een werkgever een probleem is om u aan te nemen” (zie ook figuur 6.2). De vraag is ingewikkeld en een deelnemer merkt op dat het niet

<sup>26</sup> Het model voor Werkverkenner 2.0 bestaat uit 18 voorspellende factoren voor werkhervatting (zie hoofdstuk 3). 12 factoren worden vastgesteld met 23 vragen aan de klant. De andere zes factoren zijn vastgelegd in de registratiesystemen van UWV.

erg is om over een antwoord na te moeten denken, maar dat de vraag wel meteen duidelijk moet zijn.

Geef aan of de situatie voor een werkgever een probleem is om u aan te nemen.

Uw financiële situatie

1 punt: Helemaal geen probleem

5 punten: Zeer groot probleem

1  2  3  4  5

De opvang van uw kind(eren)

1 punt: Helemaal geen probleem

5 punten: Zeer groot probleem

1  2  3  4  5

Figuur 6.2: Voorbeeld van een opvallend geformuleerde vraag volgens de deelnemers.

De vormgeving van de UWV website wordt als goed tot voldoende beoordeeld (criterium 5). De UWV huisstijl ziet er volgens de deelnemers betrouwbaar en professioneel uit. De deelnemers weten goed waar ze zich in de vragenlijst bevinden. Hyperlinks en knoppen worden duidelijk herkend. Het merendeel van de deelnemers maakt zich niet zo druk over privacy issues (criterium 6). De Werkverkenner is onderdeel van de werkmap en is beveiligd via Digi-D. Dit wordt als veilig ervaren en men heeft hier voldoende vertrouwen in. Ook noemt een deelnemer dat UWV ergens heeft genoemd dat de gegevens vertrouwelijk worden behandeld. UWV wordt ook gezien als een organisatie die te vertrouwen is.

#### 6.2.1.2 Conclusies en aanbevelingen

De deelnemers waren over het algemeen positief over de vragenlijst, de gebruiksvriendelijkheid en over de vragen zelf. Alle deelnemers beoordeelden alle onderdelen (m.u.v. één deelnemer op één onderdeel) met een voldoende of een goed. Op basis van de resultaten kunnen we concluderen dat de Werkverkenner door de deelnemers als gebruiksvriendelijk wordt ervaren.

#### Aanbevelingen inhoud en doel

Tijdens de user-walkthrough kwam naar voren dat het voor deelnemers soms niet duidelijk is waarom zij de Werkverkenner moeten invullen, wat het nut voor henzelf is en wat ze kunnen verwachten van UWV. Een paar bevindingen:

- > De *reden* waarom de UWV Werkverkenner ingevuld moet worden en wat het *nut* voor UWV en WW-klant zelf is, is voor sommige respondenten niet of voor anderen onvoldoende duidelijk. De deelnemers hebben het gevoel dat ze de vragenlijst meer voor het UWV invullen om een beeld te geven van zichzelf in relatie tot werk dan voor zichzelf. Zij vragen zich af wat het UWV doet met deze informatie en in hoeverre deze informatie van invloed is op verdere acties van UWV. De deelnemers zien het, bijvoorbeeld, niet terug in de vorm van coaching of tips op maat.
- > De houding ten opzichte van de vragenlijst is wisselend. Sommigen zijn bang voor ongewenste interpretaties van de antwoorden waar op gestuurd gaat worden door het UWV. Anderen noemen de betrouwbaarheid van de uitkomsten als je de vragenlijst wel/niet eerlijk invult<sup>27</sup>. Sommigen zien de Werkverkenner als een taak in hun takenlijst die moet gebeuren en anderen zien het als iets vrijblijvends.

<sup>27</sup> Er zijn analyses verricht naar sociaal wenselijke beantwoording van de vragen. Zie hiervoor hoofdstuk 7. Deze analyses betroffen het inspecteren van de verdelingen van de antwoorden op de items. De resultaten van deze analyses geven geen indicatie voor een hoge mate van sociaalwenselijke beantwoording.

- > De verwachtingen van het proces dat een WW-klant doorloopt liepen erg uiteen. Voor sommige deelnemers was het direct duidelijk dat de meeste communicatie via de werkmap verloopt en het eerste persoonlijke contact plaatsvindt na vier maanden<sup>28</sup>. Een andere deelnemer dacht dat hij naar aanleiding van het invullen van de vragenlijsten een gesprek zou krijgen met een persoonlijke coach bij zijn sollicitatie activiteiten en het vinden van werk (persoonlijk advies).

Op verschillende manieren kan aan bovenstaande verbeterd worden:

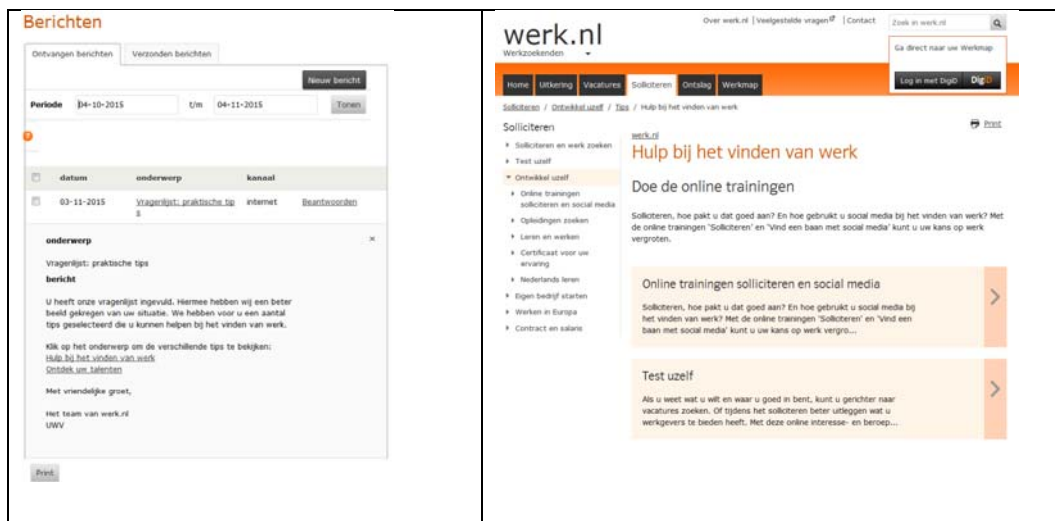
- > Spreek de persoon persoonlijk aan. Geef een positieve uitleg (bijvoorbeeld via een videoboodschap in plaats van alleen tekstueel), waarin naast informatie en instructies ook empathie wordt getoond (bijvoorbeeld 'Wat fijn dat u meedoet', 'Er zijn geen goede of foute antwoorden, het gaat om uw mening', 'Misschien vraagt u zich af waarom bepaalde vragen gesteld worden, maar dit doet het UWV omdat..., etc. ).
- > Zorg dat de instructie de volgende onderwerpen bevat: doel van de vragenlijst (reden en nut van de vragenlijst voor het UWV en de WW-klant). Wat doet UWV met de gegevens?.
- > Geef aan dat de beleving en mening van de WW-klant belangrijk is en dat er bij een aantal vragen geen goede of foute antwoorden zijn.)
- > Uitleg over het proces en vervolgstappen (schep geen valse verwachtingen over persoonlijk versus digitaal contact; probeer hier begrip bij de WW-klanten te krijgen). Hier kan in de rapportage na het invullen van de vragenlijst uitgebreider op in worden ingegaan.
- > Omgang met privacy en persoonlijke gegevens.
- > Geruststellen dat er geen tijdsdruk is en dat de antwoorden aangepast kunnen worden.
- > Uitleg van de functies op het scherm (wellicht in een aparte video of onder de help knop waar in deze video naar wordt verwezen).

#### *Aanbevelingen feedback, tips en rapportage*

Na het invullen van de Werkverkenner krijgen de WW-klanten een samenvatting van hun antwoorden. Vrij snel hierna krijgen deelnemers een of meerdere geautomatiseerde tips (uit een beperkte set van ongeveer tien standaard tips) in hun Werkmap, afhankelijk van hun scores op de factoren van de vragenlijst (zie figuur 6.3). Denk hierbij aan tips over leren over solliciteren (online training en sociale media), erachter komen wat je kunt (test uzelf, eigen bedrijf starten, ontwikkel uzelf), hulp bij vinden van werk.

---

<sup>28</sup> Dit was het dienstverleningsmodel in de periode dat de WW-klanten deelnamen aan de user-walkthrough. In principe kregen de meeste WW'ers voornamelijk online dienstverlening en in de vierde maand een persoonlijk gesprek met een adviseur van UWV. In het najaar van 2016 is UWV gestart met een nieuw WW-dienstverleningsmodel. De nieuwe dienstverlening combineert de algemene online dienstverlening met een meer persoonlijke benadering. Wanneer op basis van de Werkverkenner 1.0 blijkt dat een werkzoekende een (zeer) lage kans op heeft, dan wordt hij/zij snel uitgenodigd voor een persoonlijk gesprek. De werkzoekenden met een hoge kans op werk worden rond een half jaar uitgenodigd voor een persoonlijk gesprek, als blijkt dat zij nog geen werk gevonden hebben (zie ook het UWV jaarverslag 2016).



Figuur 6.3: Tips na het invullen van de Werkverkenner.

Met betrekking tot de rapportage en de tips online na invullen van de Werkverkenner waren de reacties wisselend. Over het algemeen worden de tips als duidelijk en kernachtig ervaren en de WW-klienten halen eruit wat voor hen relevant is. Echter, slechts een enkeling had concreet iets aan de tips. Het merendeel van de deelnemers ervaart de informatie als reeds bekend en te algemeen. Een aantal deelnemers ervaart dat iedereen dezelfde standaard tips krijgt. Het aanbod van de informatie is bovendien beperkt (wisselt niet). Er is behoefte aan een rapportage met informatie en kennis op maat en bij voorkeur gekoppeld aan vacatures. Niet alle deelnemers blijken op de hoogte te zijn van andere vormen van tips en feedback, zoals het volgen van webinars en online cursussen. Hierbij kan men feedback krijgen op persoonlijke vragen (Hoe ga je om met een 'gat' op je CV? of 'verkeerde werkervaring', solliciteren via sociale media) van cursisten en werk adviseurs. Het bekijken van de video's is niet verplicht. Gezien de bevindingen uit de user-walkthrough dat de deelnemers behoefte hebben aan een rapportage die beter aansluit bij hun situatie, worden hieronder worden hiervoor aanbevelingen gedaan:

- > Zorg dat de rapportage en tips zoveel mogelijk aansluiten bij de specifieke situatie/scores van de klant.
- > Probeer iets te veranderen aan de beleving van de werkzoekende, welke nu nog onvoldoende ervaart dat de tips gericht zijn op de persoonlijke situatie en scores. Dit kan bijvoorbeeld door de verwoording of presentatie van de tips en terugkoppeling aan te passen.
- > De rapportage sluit beter aan bij de situatie WW-klienten als er tips worden gegeven met concrete suggesties voor stappen die men kan zetten, zoals een koppeling naar actuele vacatures. Denk hierbij goed na over welke informatie en stijl van feedback mensen motiveert en helpt bij het vinden van werk (Abraham & Michie, 2008; Blanson Henkemans, 2009; Paulissen et al., 2013; Hermsen et al., 2014).

#### *Aanbevelingen vervolgonderzoek onder gebruikers werkzoekenden*

De belangrijkste aanbevelingen hierboven hebben betrekking op (extra) informatie voor de WW-klient over de achtergrond voor het stellen van bepaalde vragen en de manier waarop de resultaten verwerkt/besproken worden. Een andere algemene aanbeveling is om de Werkverkenner ook te testen bij andere doelgroepen, die niet tijdens deze user-walkthrough zijn meegenomen. Denk aan bijvoorbeeld lager opgeleiden, mensen met een niet-Nederlandse nationaliteit en minder digivaardigen. Het zou kunnen dat voor deze groepen

meer ondersteuning nodig is, zoals minder informatie op een pagina. Meer onderzoek zal dit moeten uitwijzen.

De user-walkthrough is een vorm van kwalitatief onderzoek waarbij, in dit geval, met een kleine groep WW-klienten uitgebreid is ingegaan op de ervaringen met de Werkverkenner. Overwogen kan worden aanpassingen naar aanleiding van de user-walkthrough in een grotere, representatieve groep respondenten te testen. Dit zou, bijvoorbeeld, gedaan kunnen worden met een feedback vragenlijst na het aanpassen van de introductie of tips op maat.

## 6.2.2 Gebruikersonderzoek onder adviseurs werk

### 6.2.2.1 Onderzoeksopzet

Het doel van de focusgroep sessie was om de adviseurs actief mee te laten denken over de doorontwikkeling van de Werkverkenner. Daarbij lag de focus op de gebruiksvriendelijkheid van het instrument en hoe het instrument zo goed en verantwoord mogelijk benut kan worden in de praktijk. Onderwerpen die aan bod komen zijn:

- > Inventarisatie huidige situatie (gebruik van de resultaten van de Werkverkenner 1.0 in het UWV registratiesysteem) en gebruik Werkverkenner in het (huidige en toekomstige) dagelijks werk: hoe kan de rapportage van de Werkverkenner de adviseurs werk zo goed mogelijk ondersteunen in het voeren van een gesprek met de WW-klant?;
- > Feedback op o.a. begrijpelijkheid, gebruik in de praktijk m.b.t. diverse vormen van feedback, huidige en ideale toekomstige scenario's/rapportages;
- > Feedback op de handleiding;
- > Inventarisatie van positieve punten, aandachtspunten en oplossingsrichtingen.

UWV heeft 12 professionals die werkzaam zijn in verschillende regio's (van Heerlen tot Amsterdam) uitgenodigd voor de focusgroep sessie op 15 september 2016 op het hoofdkantoor van UWV in Amsterdam. Dit zijn:

- > 4 adviseurs basisdienstverlening
- > 3 adviseurs intensieve dienstverlening
- > 3 adviseurs kwaliteit en implementatie,
- > 1 productontwikkelaar van de Werkverkenner,
- > 1 adviseur leren en ontwikkelen (UWV Academie)

De opbouw van de sessie was als volgt:

*Huidige versie - Werkverkenner 1.0:* Tijdens de eerste sessie inventariseerde de 'facilitator' de ervaringen van de adviseurs met het gebruik van de Werkverkenner in het dagelijks werk (hoe ondersteunt de Werkverkenner de adviseurs in hun werk en wat zijn verbeterpunten?). De focus ligt hierbij op gebruiksgemak en niet op het proces. De deelnemers gaven feedback op huidige versie van de Werkverkenner en van de rapportage die zij te zien krijgen in UWV registratiesysteem (SONAR), waarbij ze tevens aangaven hoe hun ideale rapportage eruit zou zien.

*Mogelijke toekomstige rapportage - Werkverkenner 2.0:* Tijdens de tweede interactieve sessie werd door TNO de stand van zaken weergegeven van het onderzoek naar de doorontwikkeling en werden drie nieuwe mogelijke vormen van rapportages (i.e. toekomstscenario's) gepresenteerd, gebaseerd op de tussenresultaten van het onderzoek. De deelnemers gaven tijdens de sessie feedback op deze toekomstscenario's. De scenario's werden aan de hand van meerdere werkelijke cases aangeboden. Daarnaast werden vragen gesteld over *behoefte aan training en een handleiding*.

## 6.2.2.2 Bevindingen

### *Input medewerkers UWV op huidige versie – Werkverkenner 1.0*

Figuur 6.4 laat zien hoe de huidige terugkoppeling van de resultaten van de Werkverkenner (voorspelling kans op werk en diagnose) van een klant eruit ziet in het systeem dat de adviseurs gebruiken (SONAR). De adviseurs kunnen dit profiel raadplegen en gebruiken als voorbereiding op hun gesprek met de WW-klant.

DV	Soort profiel	Totale score	Aangemaakt c	Aangemaakt door
> DV-20151216	Werkverkenner 1.0	67,31	17-12-2015 23:02:1	SADMIN

Naam Profiel kenmerk	Waarde	Score	Diagnose	Indicatie
> Leeftijd		29,79		<input type="checkbox"/>
Gewerkte jaren in laatste functie		80	Belemmerend	<input type="checkbox"/>
Problemen met de Nederlandse taal		0		<input type="checkbox"/>
Lezen		2		<input type="checkbox"/>
Schrijven		2		<input type="checkbox"/>
Luisteren		2		<input type="checkbox"/>
Spreken		2		<input type="checkbox"/>
Visie terugkeer op werk		41,67	Belemmerend	<input type="checkbox"/>
Zich te ziek voelen om te werken		0		<input type="checkbox"/>
Actief zoekgedrag: direct contact w		0	Belemmerend	<input type="checkbox"/>
Werkzoekintentie		75		<input type="checkbox"/>
Externe variabele attributie		61,11	Belemmerend	<input type="checkbox"/>
Algemeen werkvermogen		100		<input type="checkbox"/>
Lichamelijk werkvermogen		100		<input type="checkbox"/>
Psychisch werkvermogen		100		<input type="checkbox"/>

Figuur 6.4: Weergave van de huidige rapportage van de resultaten van de Werkverkenner 1.0 in Sonar betreffende één WW-klant. Totale score = kans op werk binnen 12 maanden.

De adviseurs gaven de volgende input op de huidige versie van de rapportage van de Werkverkenner in SONAR

*Behoeftte aan taakondersteuning:* De deelnemers geven aan dat de huidige Werkverkenner inzicht geeft in de situatie van de WW-klant en goede gesprekpunten biedt. Het kan als voorbereidingsinstrument ingezet worden. Soms vinden de adviseurs het lastig hoe ze kunnen doorvragen tijdens het gesprek met de werkzoekende aan de hand van de resultaten van de Werkverkenner. Zij zouden het fijn vinden om voorbeeldvragen te kunnen inzien (of eventueel als onderdeel van de opleiding). Een aantal adviseurs geven als aanbeveling mee dat er in de training of introductie van het instrument meer aandacht besteed mag worden aan de theoretische onderbouwing en het onderzoek naar het instrument.

*Behoeftte aan alle informatie uit verschillende bronnen in één rapportage/systeem:* Diverse adviseurs werk geven aan dat er wat betreft de WW-klant veel verschillende documenten

zijn. Dit maakt het voor de adviseurs onoverzichtelijk. Het zou fijn zijn om de informatie vanuit andere bronnen gecombineerd aangeboden te krijgen (of met elkaar verbonden in de vorm van linken). Andere bronnen die genoemd zijn betreffen: (1) de CV-kaart; (2) de beroepenkaart; (3) lokale arbeidsmarktinformatie; (4) de keuzehulp dienstverlening. Deze laatste ondersteunt de adviseur in de keuze voor het in te zetten relevante (en vaak bewezen effectieve) dienstverlening.

*Terugkoppeling van de resultaten:* Ten aanzien van de terugkoppeling van de uitkomsten van de Werkverkenner in Sonar vindt men het belangrijk dat dit in één oogopslag in te zien is. Dat het veld niet onderbroken is en men naar beneden moet scrollen. Dat er de mogelijkheid bestaat tot meer informatie (door uitklapbaarheid van de informatie; of mogelijkheid tot het opvragen van meer informatie door op een "i"-button te kunnen klikken, of opties van vervolgacties te zien door op een "a"-button te kunnen klikken). Het is momenteel niet in een oogopslag duidelijk welke vragen bij welke factor horen. Het zou fijn zijn om bij de factoren (bijvoorbeeld door er met de muis op te gaan staan) de achterliggende vragen te zien. Met betrekking tot de factoren waarop de werkzoekende belemmerend scoort zouden de professionals het ook handig vinden om niet alleen de achterliggende vragen, maar ook de antwoorden van de werkzoekende te kunnen inzien (wederom bijvoorbeeld door het scherm uit te kunnen klappen, door te kunnen klikken, of boven een factor de muis te positioneren).

*Meer kleuren en behoefte aan inzien achterliggende vragen:* Momenteel maakt de terugkoppeling in SONAR slechts gebruik van twee kleuren (groen en rood). Meer nuance zou fijn zijn. Genoemde aanbeveling is: breng meer nuance aan in de kleuren, zoals oranje. De adviseurs zouden het fijn vinden om op dezelfde pagina in SONAR niet alleen de uitslagen op de factoren, maar ook de achterliggende vragen en het ingevulde antwoord door WW-klant te kunnen inzien te kunnen inzien. Nu staan deze op verschillende plaatsen. De vragen en ingevulde antwoorden zouden toegankelijk gemaakt kunnen worden achter het 'stoplicht'.

*Volgorde van de factoren:* De volgorde waarin de huidige factoren worden aangeboden in de uitslagen in SONAR is niet logisch. Zij lijken niet op volgorde van belangrijkheid geplaatst. Problemen Nederlandse taal staat nu vrij prominent bovenaan, terwijl dit slechts voor een zeer beperkt deel van de werkzoekenden een belemmering vormt. Bovendien wordt deze factor vervolgens opgedeeld in lezen, schrijven, luisteren en spreken, waarmee het feitelijk 5 regels inneemt. De adviseurs vinden deze onderverdeling weinig toevoegen.

*Weglaten van de cijfers* (kolom: Score). De professionals weten niet goed wat de cijfers zijn die ze bij iedere factor tegenkomen. Zij raden aan om deze weg te halen, minder prominent in beeld te zetten of hier duidelijkheid over te geven wat dit betekent en hoe zij deze kunnen lezen.

*Interpretatie van de resultaten en betekenis factoren:* Het begrip over de relatie tussen de kans op werkherhaling (totale score in huidige SONAR) en de scores op de individuele factoren (de eigenlijke snelle diagnose van het instrument Werkverkenner) zou bij sommige professionals verbeterd kunnen worden. Dit zou bijvoorbeeld kunnen door aanpassingen in achterliggende informatie in de terugkoppeling van de uitslagen van de Werkverkenner in SONAR. Een andere/aanvullende mogelijkheid is dit mee te nemen in de opleiding van adviseurs werk. Niet voor alle adviseur is direct duidelijk wat 'totaal score' betekent en de 'kans op werk'. Ook is de betekenis van sommige factoren niet direct duidelijk (o.a. de factor Externe attributie). Het zou heel fijn zijn om meer uitleg in Sonar te krijgen met betrekking tot de factoren die belemmeringen blijken voor een werkzoekende. Genoemd zijn: uitleg over de



betekenis van de factoren in het algemeen; uitleg over hetgeen het betekent als een werkzoekende een belemmerende score heeft; inzicht in welke vragen corresponderen met welke (belemmerende) factor; inzicht in de antwoorden van de werkzoekenden op de vragen van iedere belemmerende factor.

*Feedback aan de klant zelf?* De tips en adviezen die de werkzoekende ontvangt na het invullen van de Werkverkenner (in de digitale omgeving) worden niet altijd ingezien. Het zou goed zijn om een reminder te sturen als UWV constateert dat een werkzoekende hier niets mee heeft gedaan.

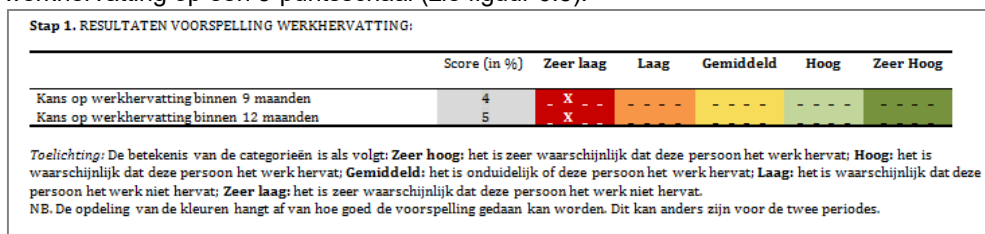
*Tonen van de kans op werkhervatting aan de WW-klant?* Er wordt gediscussieerd over het wel/niet tonen van de uitslagen van de Werkverkenner aan de werkzoekende tijdens bespreking. De huidige manier van handelen zoals dat momenteel gebeurt bij Competentie Test Centrum lijkt een goede richtlijn. Deze richtlijn komt erop neer dat de professional over het algemeen de bespreking doet aan de hand van (het tonen van) de rapportage, echter kan de professional ook hiervan afwijken. Dit is toegestaan als de professional inschat dat het tonen van de resultaten een averechts effect hebben (iemand bijvoorbeeld nog meer gedemotiveerd raakt om werk te zoeken). De professional moet wel kunnen beargumenteren waarom die deze keuze maakt. NB: een werkzoekende heeft altijd recht tot inzicht in eigen gegevens als deze erom vraagt, ongeacht of de professional inschat dat dit averechts kan werken.

#### Input medewerkers UWV op mogelijke toekomstige Werkverkenner

De toekomstscenario's zijn in drie stappen gepresenteerd: (stap 1: resultaten voorspelling werkhervatting, stap 2 resultaten achtergrondkenmerken en stap 3 Resultaten zachte factoren uit de vragenlijst.

#### Stap 1: visualiseren van de voorspelling werkhervatting:

De adviseurs staan positief tegenover de schalen met kleurcodering: de kans op werkhervatting op een 5-puntsschaal (zie figuur 6.5).

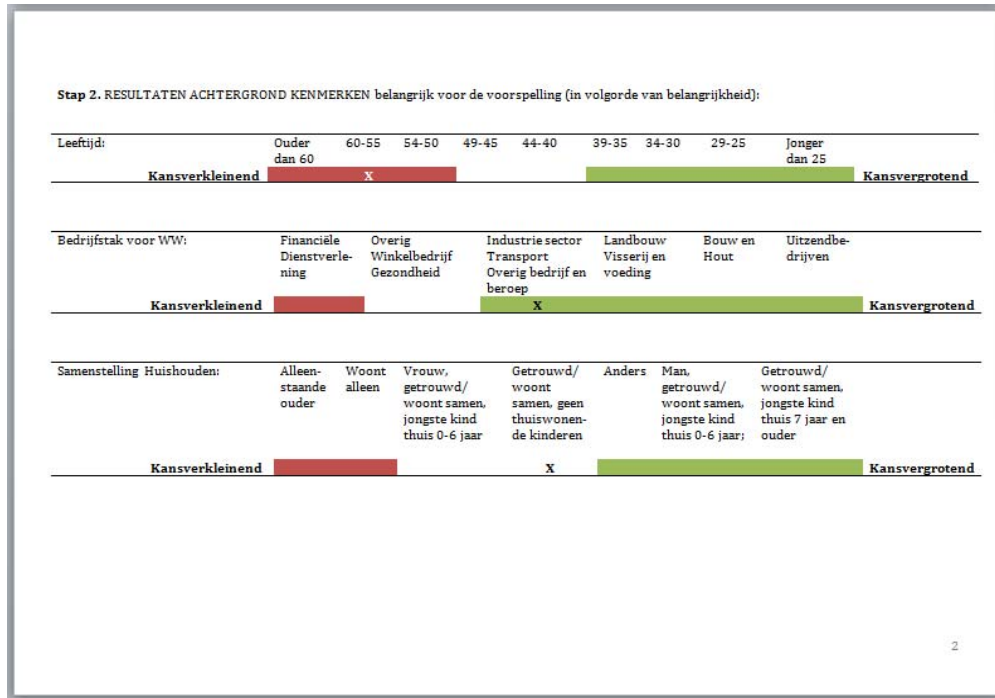


Figuur 6.5: Stap 1 Resultaten voorspelling werkhervatting met een 5-puntsschaal

Voordeel van de vijf puntsschaal is dat het ook een middengroep toont, waarvan de voorspelling minder accuraat is dan van de extremen (d.w.z. dat voor personen in deze middengroep het voorspelmodel vaker een missclassificatie maakt). Dit vindt de opdrachtgever UWV belangrijke informatie. De deelnemers vinden de vijf punts variant iets lastiger. Wat betekent in de vijf puntsschaal de kleur geel ( gemiddeld): nodig je dan mensen wel of niet uit voor een face-to-face gesprek? Hebben ze dan een kleinere kans? Dat is niet direct duidelijk. De deelnemers geven aan dat het belangrijkste is dat de indeling aansluit bij het beleid van UWV. Momenteel is dat gebaseerd op een indeling in vieren. Belangrijk blijft natuurlijk dat de keuze uiteindelijk ook tegen mogelijke veranderingen in beleid bestand moet zijn. Voor beide varianten is dus iets te zeggen.



Stap 2: Visualiseren resultaten achtergrond kenmerken

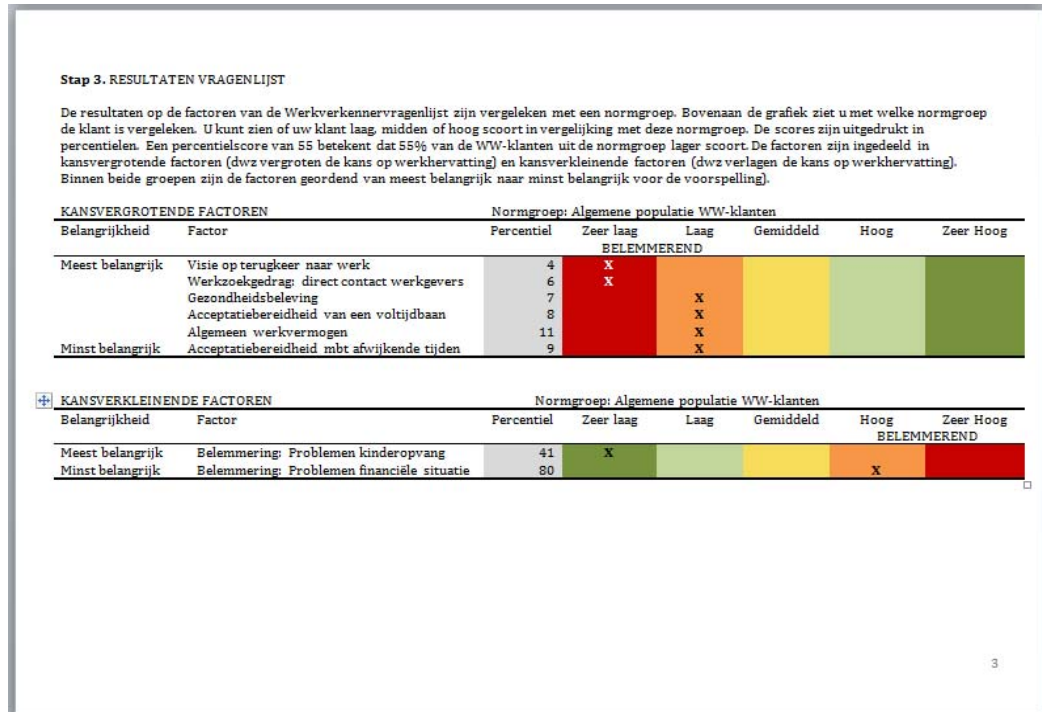


Figuur 6.5: Resultaten achtergrondkenmerken.

Noot: De scenario's zijn gebaseerd op tussenresultaten. De eindresultaten m.b.t. de normering uit par. 6.2 waren nog niet bekend. Daarom wijkt de indeling van samenstelling huishouden in dit scenario af van de definitieve vastgestelde indeling. Ook wordt in dit scenario Bedrijfstak voor WW nog gehanteerd als factor ten behoeve van diagnose. In par. 6.1 hebben we geconcludeerd om deze factor op te nemen in het voorspelmodel, maar dat UWV voor de diagnose beter gebruik kan maken van de actuele arbeidsmarktinformatie die UWV beschikbaar stelt op haar website. Voor het doel om te toetsen of de scenario's bruikbaar zijn voor de adviseur, is het niet belangrijk dat de eindresultaten nog niet bekend waren.

De deelnemers vinden het overzicht met de achtergrondkenmerken duidelijk.

Stap 3: Visualiseren resultaten vragenlijst (kansvergroten- en verkleinende factoren)



Figuur 6.6: Resultaten kans vergroten- en verkleinende factoren. De kleuren die hier gebruikt zijn wijken af van de kleuren in hoofdstuk 6 van deze eindrapportage (geel is licht groen geworden).

Noot: De scenario's zijn gebaseerd op tussenresultaten. Zoals te zien is in Stap 3 is voor de focusgroep sessie destijds gekozen voor de normgroep algemene populatie WW-klienten. Voor de adviseur zijn de resultaten van een werkzoekenden op de Werkverkenner alleen bruikbaar als hij ze kan afzetten tegen een succesvolle werkzoekende, dat wil zeggen een werkzoekenden die binnen een jaar een baan gevonden heeft. Daarom is de normering bepaald aan de hand van de normgroep: werkherhvatting binnen 12 maanden (zie par. 6.1). Voor het doel om te toetsen of de scenario's bruikbaar zijn voor de adviseur, is het niet belangrijk dat in dit voorbeeld een andere normgroep gepresenteerd is.

De deelnemers merken op dat het verwarrend kan zijn dat een 'goede uitkomst' bij kansvergrotenende factoren aan de andere kant staat dan bij kansverkleinende factoren (zie figuur 6.6). Zij geven aan dat sommige mensen kleurenblind zijn en wellicht deze verandering niet goed doorhebben. De deelnemers zien liever dat de twee tabellen dezelfde richting op zijn te lezen, dat wil zeggen: alles wat 'goed' is een kant op en alles wat 'slecht' is de andere kant op. Verder wordt opgemerkt dat er teveel labels zijn in de tabel. In de linker kolom staat "meest belangrijk" en "minst belangrijk", dat kan weg. Het is voldoende als in een toelichting wordt uitgelegd dat de factoren zijn geordend van meest naar minst belangrijk. Ook vindt men het niet nodig dat er boven de kleuren zeer laag t/m zeer hoog staat.

De percentielen genoemd onder stap 3 (in de grijze kolom) zijn een verfijning van hoe iemand scoort ten aanzien van de totale populatie WW-klienten. Echter lijken veel deelnemers hier niet direct warm van te worden. Het vergt meer uitleg/inzicht om te kunnen begrijpen; en het wordt ervaren als minder noodzakelijk omdat je al weet waar de werkzoekende op een vijf puntsschaal scoort. Er wordt geconcludeerd dat je de percentielen

misschien niet direct in het zicht te plaatsen, maar in te zien is als je erop klikt. Het is daarmee een verdieping voor diegenen die dat willen.

#### *Toelichting op de betekenis van de factoren*

Voor elke factor is een toelichting geschreven die de adviseurs kunnen gebruiken bij de interpretatie van de factor. M.b.t. de toelichting op de factoren wordt er wisselend gereageerd door de adviseurs. Sommigen vinden dit zeer verhelderend; anderen ervaren het als veel tekst. Als opmerking wordt genoemd dat de toelichting op 1 scherm moet passen, zodat je de informatie kunt 'screenen'. Allen zijn het eens dat dit heel goed is voor instructiemateriaal of naslagwerk ter voorbereiding van het gesprek met de klant. Er wordt uiteindelijk genoemd dat het fijn zou zijn als deze informatie beschikbaar is m.b.t. de factoren waarop de werkzoekende belemmerend scoort. Het aanbod van deze informatie zou wederom d.m.v. doorklikken/uitklappen/ergens op gaan staan beschikbaar moeten worden. Op deze manier krijgt men niet te veel informatie in een keer en is het mogelijk om naar behoefte te verdiepen.

#### *6.2.2.3 Conclusies en aanbeveling*

De scenario's, die de resultaten van de toekomstige Werkverkenner 2.0 weergeven aan de adviseurs werk, worden over het algemeen goed ontvangen. De voorkeur gaat uit naar meer nuancering in kleuren in de terugkoppeling van de factor dan de huidige twee kleuren rood en groen. De weergave is echter nog niet compact genoeg en de tekst eromheen moet flink worden gereduceerd. (NB. Deze aanbeveling hebben verwerkt in het nieuwe overzicht, zie Tabel 6.23 met de uitslag van de voorbeeldcasus Pierre). Verder is er behoefte aan mogelijkheden tot het bekijken van extra resultaten en informatie op verzoek (bijv. door te klikken op een i-button) en er is behoefte aan handvaten voor acties naar aanleiding van de resultaten (bijv. voorbeeldvragen of verwijzing naar e-producten). We bevelen aan om de input van de adviseurs te verwerken in de terugkoppeling (en implementatie in SONAR) van Werkverkenner 2.0 aan de adviseur.

## 7 Psychometrische eigenschappen

### Conclusies:

Op basis van de resultaten van de psychometrische analyses concluderen we dat er sprake is van een betrouwbaar en valide instrument. Dit impliceert dat de scores op de factoren die bestaan uit meerdere items van de Werkverkenner gebruikt kunnen worden voor een eerste diagnose. Deze diagnose wordt na invulling van de Werkverkenner teruggekoppeld aan de adviseur werk. De werkzoekende ontvangt op basis van de diagnose een online advies. UWV gebruikt de diagnose om te bepalen welke dienstverleningsaanbod nodig is voor wie. Omdat voor de factoren die gemeten zijn door één item slechts één van de psychometrische eigenschappen kon worden onderzocht, is enige voorzichtigheid geboden om op basis van deze losse items een diagnose te bepalen en op basis hiervan een advies te geven.

### Betrouwbaarheid

De betrouwbaarheden op basis van *inter-itemrelaties* van de factoren in de totale onderzoeksgroep en in de normgroep zijn voldoende tot hoog. Voor de subgroepen is de betrouwbaarheid van de meeste factoren voldoende tot hoog. Dit zijn belangrijke bevindingen, omdat hiermee is aangetoond dat items die samen een factor vormen sterk met elkaar samenhangen (i.e. hoge consistentie). Op basis van de waarden van de test-hertest correlaties concluderen we dat de *test-hertest betrouwbaarheid* van de meeste factoren voldoende is voor de totale WW-representatieve groep, de normgroep en de subgroepen. De test-hertest correlaties van de meeste factoren in Werkverkenner 2.0 zijn ruim voldoende ( $r \geq 0,70$ ). We kunnen concluderen dat de betrouwbaarheid voldoende is voor het doel waar de afzonderlijke factoren voor ingezet gaan worden (i.e. voor diagnose en begeleiding).

### Constructvaliditeit

De constructvaliditeit van de Werkverkenner is goed. Dit blijkt uit de *confirmatieve factoranalyse*: de veronderstelde factorstructuur past goed bij de geobserveerde data. Voor alle items werden *factorloadingen* van voldoende grootte gevonden. Daarnaast is deze factorstructuur ook houdbaar voor de normgroep en de subgroepen. Er zijn geen aanwijzingen voor *item bias*. Dit betekent dat we scores op de factoren kunnen vergelijken tussen groepen. In het *soortgenotenonderzoek* van de Werkverkenner vinden we voor enkele schalen (i.e. factoren die door meerdere items worden gemeten) samenhang met inhoudelijk corresponderende schalen uit de Arbeidsmarktmotivatie en Werkzoekgedrag Vragenlijst (AWV); met name tussen de factor Visie op terugkeer naar werk, de belangrijkste zachte factor in Werkverkenner 2.0 en de factor Inschatting kans op werk ( $r = 0,69$ ). Voor de meeste schalen vinden we echter een minder hoge samenhang dan we van tevoren hadden verwacht. De AWV sluit niet op alle gebieden aan bij de Werkverkenner. Daarnaast zijn de Werkverkenner en de AWV niet op dezelfde dag ingevuld, wat een invloed kan hebben gehad op de samenhang tussen de schalen. Op basis van het soortgenoten onderzoek is er geen reden om schalen van de Werkverkenner niet mee te nemen in het voorspelmodel voor werkhervatting.

### Sociaalwenselijke beantwoording

Om de mate van sociaal wenselijkheid te onderzoeken hebben we, gegeven de onderzoeksopzet van de huidige studie, ons beperkt tot het inspecteren van de verdelingen van de antwoorden op de items. Deze geven geen indicatie voor een hoge mate van sociaalwenselijke beantwoording. Een geringe mate van sociaalwenselijke beantwoording doet geen afbreuk aan de kwaliteit van de voorspelregel voor werkhervatting. De voorspelkracht van de regel zal niet afnemen, doordat het aannemelijk is dat de mate van sociaalwenselijkheid in de antwoorden van de huidige steekproef niet zal verschillen van de mate van sociaalwenselijkheid van toekomstige werkzoekenden.

### Predictieve validiteit

De predictieve validiteit van de Werkverkenner 2.0 is redelijk (AUC = 0,78). We kunnen met een logistisch regressiemodel met de 18 factoren van de Werkverkenner 2.0 de status werkhervatting binnen 12 maanden voor 70,3% correct voorspellen in de testset.

In dit hoofdstuk richten we ons op het beantwoorden van de derde onderzoeksvraag: “Wat zijn de psychometrische eigenschappen van de Werkverkenner?”. In hoofdstuk 6 hebben we laten zien dat op basis van scores op de factoren van de Werkverkenner een snelle diagnose gemaakt wordt van de belemmeringen en mogelijkheden van WW-klienten om werk te vinden. Voor het gebruik van de Werkverkenner als diagnose-instrument is het belangrijk om de psychometrische eigenschappen van het instrument (ook wel “test” genoemd) goed in kaart te brengen.

De volgende psychometrische eigenschappen zijn onderzocht:

- > betrouwbaarheid (inter-item relaties en test-hertest betrouwbaarheid);
- > constructvaliditeit (confirmatieve factor analyse, invariantie factor structuur en item bias, soortgenotentest);
- > sociaal wenselijke beantwoording;
- > predictieve validiteit.

Bij bepaling van de *betrouwbaarheid* van een test gaat het erom de invloed van meetfouten op de testscore te schatten. Meetfouten kunnen ontstaan door toevallige factoren die het meten kunnen verstoren, zoals bijvoorbeeld een telefoon die afgaat tijdens het lezen van een item van de vragenlijst. Het basisidee van de klassieke testtheorie is dat de geobserveerde score een optelling is van een betrouwbare score (dat wil zeggen, een ware score) en een meetfout. Een betrouwbare test heeft een kleine meetfout. Bij *construct validiteit* (ook wel begripsvaliditeit genoemd) gaat het erom te toetsen of de vragenlijst inderdaad de eigenschappen meet die worden verondersteld. Meet de test de bedoelde factoren of iets anders? Van *sociaal wenselijke beantwoording* is sprake als respondenten de tendentie hebben om antwoorden te geven die de werkelijkheid ‘mooier’ voorstellen dan zijn is. Bij het bepalen van de *predictieve validiteit* van een test gaat het om de mate waarin de test het criterium voorspelt, waarbij het criterium is gemeten op een later tijdstip dan het tijdstip van de testafname (voor de Werkverkenner is het belangrijkste criterium de status werkhervatting op 12 maanden na aanvang van de WW-uitkering).

Voor het beschrijven van de test-hertest betrouwbaarheid gebruiken we alle zachte factoren van de Werkverkenner<sup>29</sup>, hieronder weergegeven in tabel 7.1. Deze factoren werden in dit onderzoek gemeten met de vragenlijst Werkverkenner 1.0+. Voor het bepalen van de betrouwbaarheid gebaseerd op inter-item relaties en voor het bepalen van de constructvaliditeit gebruiken we een selectie van deze factoren, namelijk de tien factoren die met meerdere items zijn gemeten (zie tabel 7.1).

<sup>29</sup> De meeste harde factoren van de Werkverkenner komen uit de administratie van UWV (zie Tabel 1.1, Hoofdstuk 1). Veel van deze harde factoren van de Werkverkenner worden gebruikt om de hoogte van de uitkering te bepalen voor de WW-gerechtigde en/of komen uit de basisregistraties. (Er zijn 11 basisregistraties. Een basisregistratie is een door de overheid officieel aangewezen registratie met gegevens die door alle overheidsinstellingen verplicht worden gebruikt bij de uitvoering van publiekrechtelijke taken). Het is daarom aannemelijk dat de betrouwbaarheid en validiteit van deze factoren hoog is.

Voor het onderzoeken van de sociaal wenselijkheid van de antwoorden hebben we de verdelingen van de afzonderlijke items van de zachte factoren bekeken.

Bij het beschrijven van de predictieve validiteit gaan we uit van het beste voorspelmodel dat resulteerde uit de analyses uitgevoerd om werkhervatting te voorspellen (beschreven in hoofdstuk 3). Het beste voorspelmodel bevatte een beperkte set van 18 factoren van de Werkverkenner 1.0+. Deze 18 factoren vormen samen Werkverkenner 2.0 (zie tabel C uit de samenvatting en tabel 3.6 uit hoofdstuk 3).

Tabel 7.1: Overzicht van de zachte factoren die zijn opgenomen in de vragenlijst (Werkverkenner 1.0+). De eerste kolom geeft aan of de factor geselecteerd is voor Werkverkenner 2.0. De derde kolom geeft het aantal items per factor weer.

WV 2.0	Factoren	#	Formulering items
ja	Visie op terugkeer naar werk	3	Ik denk dat het lang zal duren voordat ik werk vind Ik ben vaak afgewezen, ik verwacht geen werk meer te vinden Ik verwacht snel zelf een baan te vinden
ja	Werkzoekgedrag m.b.t. direct contact werkgevers	3	Ik heb sollicitatiegesprekken gevoerd Ik heb met werkgevers gebeld Ik heb werkgevers bezocht
ja	Gezondheidsbeleving	4	Hoe beoordeelt u uw werkvermogen op dit moment, als u kijkt naar de psychische (geestelijke) eisen die werk aan u zou stellen? Hoe beoordeelt u uw werkvermogen op dit moment, als u kijkt naar de lichamelijke eisen die werk aan u zou stellen? Wat vindt u, over het algemeen, van uw gezondheid? Ik voel me te ziek om te werken
ja	Algemeen werkvermogen <sup>1</sup>	1	Stel: u geeft 10 punten aan uw werkvermogen in de beste periode van uw leven. Hoeveel punten zou u dan geven aan uw werkvermogen (1 t/m 10)?
ja	Aantal uren in staat te werken <sup>2</sup>	1	Hoeveel uren per week bent u in staat om te werken?
ja	Balans voor- en nadelen van niet werken <sup>2</sup>	1	Werkloosheid heeft voordelen en nadelen. Vindt u dat niet-werken meer voordelen heeft dan nadelen?
ja	Acceptatiebereidheid m.b.t. afwijkende tijden	2	In hoeverre wilt u werken op onregelmatige werktijden (werken op verschillende tijden)? In hoeverre wilt u buiten kantoor tijden werken ('s avonds en 's nachts)?
ja	Acceptatiebereidheid m.b.t. voltijd baan <sup>2</sup>	1	Wilt u werken in een voltijd baan (35-40 uur per week)?
nee	Acceptatiebereidheid m.b.t. onwenselijk werk	2	Wilt u werk onder uw niveau? Wilt u werk voor een lager salaris?
nee	Financiële noodzaak <sup>2</sup>	1	Hoe belangrijk is het voor u, financieel gezien, om binnen twee maanden een betaalde baan te vinden?
nee	Belemmering: financiële problemen <sup>2</sup>	1	Is uw financiële situatie voor een werkgever een probleem om u aan te nemen?
nee	Belemmering: gebrek kinderopvang <sup>2</sup>	1	Is de opvang van uw kind(eren) een probleem voor een werkgever om u aan te nemen? (alleen bij thuiswonende kinderen)

WW 2.0	Factoren	#	Formulering items
nee	Subjectieve norm familie en partner <sup>3</sup>	1	Hoe belangrijk vinden uw familie en partner dat u betaald werk zoekt? Hoe belangrijk is de mening van deze personen voor u?
nee	Self-efficacy m.b.t. sollicitatievoorbereiding	3	Ik kan goed informatie vinden over vacatures Ik kan mijn sterke en zwakke punten voor een baan uitleggen Ik kan goed een (digitale) sollicitatie schrijven
nee	Werkzoekintentie	3	Ik ben van plan om er de komende maand alles aan te doen om werk te vinden Ik weet hoe ik de komende maand kan proberen werk te vinden Ik ben zeer gemotiveerd om komende maand werk te zoeken
nee	Werkzoekattitude m.b.t voordeel en aangenaamheid	2	Wat vindt u van het zoeken naar een betaalde baan, prettig of onprettig? Wat vindt u van het zoeken naar een betaalde baan, nadelig of voordelig?
nee	Werkzoekattitude m.b.t. nut en noodzaak	2	Wat vindt u van het zoeken naar een betaalde baan, onnodig of nodig? Wat vindt u van het zoeken naar een betaalde baan, nutteloos of nuttig?
nee	Externe variabele attributie	3	Ik denk dat je voor het krijgen van een baan vooral geluk moet hebben Ik denk dat je voor het krijgen van een baan vooral op het juiste moment op de juiste plaats moet zijn Ik denk dat de meeste mensen een baan vinden door de invloed van anderen

<sup>1</sup>Het item behorend bij deze factor wijkt qua antwoordschaal af van de andere items die hetzelfde meten.

<sup>2</sup>Het gaat hier om 1 geobserveerd item. Er is dus geen sprake van een achterliggende factor.

<sup>3</sup>De twee items behorende bij de factor subjectieve norm worden vermenigvuldigd om tot één score te komen. De items worden niet opgevat als twee indicatoren van een achterliggende factor.

## 7.1 Betrouwbaarheid

**Conclusie:** De *betrouwbaarheden gebaseerd op inter-item relaties* (Cronbach's alfa en Guttman's lambda<sup>2</sup>) van de tien factoren die gemeten zijn door meerdere items zijn voldoende tot hoog voor het doel waarvoor de factoren van de Werkverkenner ingezet gaan worden (i.e. diagnose en begeleiding; minimaal vereiste betrouwbaarheid is 0,70). Dit geldt voor de totale onderzoeksgroep, de normgroep (i.e. de groep werkhervatters) en de groep niet-werkhervatters. Op basis van de waarden van Guttman's lambda<sup>2</sup> voor de subgroepen (gebaseerd op leeftijd, nationaliteit, problemen Nederlandse taal, geslacht en WIA 35-min) concluderen we dat de betrouwbaarheid van de meeste factoren voldoende tot hoog is.

Op basis van de *betrouwbaarheden gebaseerd op test-hertest correlaties* concluderen we dat de test-hertest betrouwbaarheid van de meeste factoren voldoende is voor de totale WW-representatieve groep, de normgroep en de subgroepen. De test-hertest correlaties van de meeste factoren in Werkverkenner 2.0 zijn ruim voldoende ( $r \geq 0,70$ ).



Bij bepaling van de *betrouwbaarheid* van een test gaat het erom de invloed van meetfouten op de testscore te schatten. Meetfouten kunnen ontstaan door toevallige factoren die het meten kunnen verstoren, zoals bijvoorbeeld een telefoon die afgaat tijdens het lezen van een item van de vragenlijst. Het basisidee van de klassieke testtheorie is dat de geobserveerde score een optelling is van een betrouwbare score (dat wil zeggen, een ware score) en een meetfout. Een betrouwbare test heeft een kleine meetfout.

Er zijn verschillende aspecten van betrouwbaarheid, die zich ieder richten op verschillende mogelijke “bronnen” van de meetfout. We richten ons in dit rapport op de volgende betrouwbaarheidsmaten:

- 1) Maten van betrouwbaarheid gebaseerd op inter-item relaties. Deze maten bepalen in hoeverre de items die bij een factor horen met elkaar samenhangen. Als de antwoorden op een groep items sterk met elkaar samenhangen, dan is de meetfout klein en kunnen we spreken van een betrouwbare factor.
- 2) Test-hertest betrouwbaarheid. Deze betrouwbaarheidsmaat geeft aan in hoeverre de test dezelfde scores oplevert bij herhaald meten. Een manier om dit te onderzoeken is door de test-hertest correlaties te berekenen. De test-hertest betrouwbaarheid is bepaald op een deel van het volledige bestand: een aantal respondenten in dit bestand kreeg de Werkverkenner nogmaals aangeboden.

#### 7.1.1 *Maten van betrouwbaarheid gebaseerd op inter-item relaties*

De meest gebruikte maat om de betrouwbaarheid gebaseerd op de samenhang tussen items (i.e. de inter-item relaties) weer te geven is Cronbach's alfa. Cronbach's alfa geeft een schatting van de ondergrens van de betrouwbaarheid weer. Waarden van Cronbach's alfa liggen tussen 0 en 1, waarbij een hoge waarde wijst op een hoge (goede) betrouwbaarheid. Als vuistregel geldt dat een betrouwbare factor een alfa van minimaal 0,70 moet hebben (George & Mallory, 2003). Guttman's lambda2 is een andere maat voor het weergeven van de betrouwbaarheid. In het algemeen geldt dat het verschil met Cronbach's alfa klein is en dat lambda2 altijd minstens zo groot is als Cronbach's alfa (Sijtsma, 2009). Tevens wordt verondersteld dat lambda2 dichter bij de werkelijke betrouwbaarheid ligt dan Cronbach's alfa (Sijtsma, 2009).

Daarnaast is voor de totale groep gekeken naar de item-rest correlaties per factor. Het gaat hier om correlaties tussen een item en de totaalscore op de overige items van dezelfde factor. De hoogte van de correlatie geeft aan in hoeverre het betreffende item hetzelfde meet als de andere items behorende bij de factor. Een item-rest correlatie moet tenminste 0,30 zijn voor een goede beoordeling.

Deze betrouwbaarheidsmaten konden uiteraard alleen berekend worden voor de 10 factoren van tabel 7.1 die uit meerdere items bestaan.

##### 7.1.1.1 *Resultaten Cronbach's alfa en Guttman's lambda2 voor totale onderzoeksgroep*

**Conclusie:** De betrouwbaarheden (Cronbach's alfa en Guttman's lambda2) van de factoren in de totale onderzoeksgroep zijn voldoende tot hoog voor het doel waarvoor de factoren van de Werkverkenner ingezet gaan worden (i.e. diagnose en begeleiding; minimaal vereiste betrouwbaarheid is 0,70).

De waarden van Cronbach's alfa en Guttman's lambda2 zijn berekend voor de gehele onderzoeksgroep (N = 53079) weergegeven in tabel 7.2. Deze onderzoeksgroep is representatief voor de gehele WW-populatie die de Werkverkenner krijgt aangeboden (zie hoofdstuk 5). Zoals is te zien verschillen de waarden van alfa en lambda2 nauwelijks. De



waarden van alfa en lambda2 variëren tussen 0,70 en 0,87. Alle item-rest correlaties liggen  $\geq 0,45$  (zie bijlage I).

Tabel 7.2: Waarden van Cronbach's alfa en Guttman's lambda2 per factor voor de totale onderzoeksgroep  
*N* = 53.079.

WV 2.0	Factor	Cronbach's alfa	Guttman's lambda2
ja	Visie op terugkeer naar werk	0,80	0,79
ja	Werkzoekgedrag m.b.t. direct contact werkgevers	0,72	0,72
ja	Gezondheidsbeleving	0,85	0,86
ja	Acceptatiebereidheid m.b.t. afwijkende tijden	0,87	0,87
nee	Acceptatiebereidheid m.b.t. onwenselijk werk	0,70	0,70
nee	Werkzoekintentie	0,80	0,80
nee	Werkzoekattitude m.b.t. nut en noodzaak	0,80	0,80
nee	Werkzoekattitude m.b.t. voordeel en aangenaamheid	0,77	0,77
nee	Externe variabele attributie	0,70	0,70
nee	Self-efficacy m.b.t. sollicitatievoorbereiding	0,76	0,76

#### 7.1.1.2 Resultaten Cronbach's alfa en Guttman's lambda2 voor de normgroep

**Conclusie:** De betrouwbaarheden van de factoren in de normgroep zijn voldoende tot hoog voor het doel waarvoor de factoren van de Werkverkenner ingezet gaat worden (i.e. diagnose en begeleiding; minimaal vereiste betrouwbaarheid is 0,70).

Als normgroep namen we degenen die het werk hebben hervat binnen 12 maanden (zie hoofdstuk 6). Om inzicht te krijgen in de betrouwbaarheden van de schalen voor deze normgroep hebben we de analyses apart voor deze groep gedaan en apart voor de groep niet-werkhervatters binnen 12 maanden (tabel 7.3).

Tabel 7.3: Waarden van Cronbach's alfa en Guttman's lambda2 per factor uitgesplitst naar de status werkhervatting binnen 12 maanden

WV 2.0	Factor	Werkhervatters (n = 27670)		Niet-werkhervatters (n = 25409)	
		Cronbach's alfa	Guttman's lambda2	Cronbach's s alfa	Guttman's lambda2
ja	Visie op terugkeer naar werk	0,76	0,76	0,80	0,80
ja	Werkzoekgedrag m.b.t. direct contact werkgevers	0,70	0,70	0,71	0,71
ja	Gezondheidsbeleving	0,80	0,81	0,87	0,87
ja	Acceptatiebereidheid m.b.t. afwijkende tijden	0,87	0,87	0,86	0,86
nee	Acceptatiebereidheid m.b.t. onwenselijk werk	0,70	0,70	0,71	0,71
nee	Werkzoekintentie	0,80	0,80	0,80	0,80
nee	Werkzoekattitude m.b.t. nut en noodzaak	0,77	0,77	0,83	0,83
nee	Werkzoekattitude m.b.t. voordeel en aangenaamheid	0,77	0,77	0,77	0,77
nee	Externe variabele attributie	0,70	0,70	0,69	0,69
nee	Self-efficacy m.b.t. sollicitatievoorbereiding	0,75	0,75	0,76	0,76

De waarden van Cronbach's alfa en Guttman's lambda2 variëren tussen 0,70 en 0,87 voor de normgroep werkhervatters binnen 12 maanden en tussen de 0,69 en 0,87 voor de groep niet-werkhervatters binnen 12 maanden. Alle item-rest correlaties liggen  $\geq 0,45$  (zie bijlage I).

Over het algemeen zijn de waardes van de betrouwbaarheidsmaten voor de normgroep werkhervatters iets lager dan of gelijk aan de waardes voor de niet-werkhervatters met uitzondering van de factoren Acceptatiebereidheid m.b.t. afwijkende tijden en Externe variabele attributie (zie tabel 7.2). Ook in vergelijking met de waardes van de totale onderzoeksgroep zijn de waardes van de betrouwbaarheidsmaten voor de normgroep iets lager of hetzelfde (Tabellen 7.2 en 7.3). De waardes voor de normgroep komen echter niet onder de grenswaarde van 0,70.

#### 7.1.1.3 Resultaten Guttman's lambda2 voor subgroepen op basis van achtergrondkenmerken

**Conclusie:** Op basis van de waarden van Guttman's lambda2 concluderen we dat de betrouwbaarheid van de meeste factoren voldoende tot hoog is voor de subgroepen. We concluderen dat de waarden van de factoren hoog genoeg zijn voor het doel waar de factoren van de Werkverkenner voor ingezet gaat worden (i.e. diagnose en begeleiding).

In de voorgaande paragrafen hebben we geconcludeerd dat de betrouwbaarheid van de factoren voldoende hoog is voor de totale groep en voor de normgroep. Nu willen we ook onderzoeken of dit geldt voor de volgende subgroepen gedefinieerd aan de hand van achtergrondkenmerken (i.e. geslacht, opleiding, etniciteit, leeftijd<sup>30</sup> en het hebben van een

<sup>30</sup> Werkzoekenden zijn aan de hand van leeftijd ingedeeld in de klantgroepen 50 t/m 54 jaar en 55 jaar en ouder.

WIA 35-min-indicatie<sup>31</sup>). Omdat de factor etniciteit als zodanig niet gemeten is met de Werkverkenner, maken we gebruik van beschikbare factoren die etniciteit benaderen (zogenaamde “proxy”-maten). Dit zijn de factoren Nationaliteit en Problemen met de Nederlandse taal. Voor de factor Nationaliteit maken we onderscheid tussen de groep “Nederlandse nationaliteit” (d.w.z., 1<sup>e</sup> nationaliteit Nederlands en geen 2<sup>e</sup> nationaliteit of 2<sup>e</sup> nationaliteit westers) en de groep “niet-Nederlandse nationaliteit”. Onder deze laatste groep vallen de volgende categorieën: “1<sup>e</sup> of 2<sup>e</sup> nationaliteit Pools, 1<sup>e</sup> of 2<sup>e</sup> nationaliteit niet-westers, of 1<sup>e</sup> nationaliteit westers (anders dan Nederlands en Poolse”. Omdat de percentages in deze categorieën erg klein zijn (zie H2, Tabel 2.1) hebben we ze samengenomen tot “niet-Nederlandse nationaliteit”.

1. Geslacht: mannen versus vrouwen;
2. Opleiding: laag versus middel/hoog opgeleid;
3. Nationaliteit: Nederlandse versus niet-Nederlandse nationaliteit;
4. Problemen met de Nederlandse taal: geen problemen met de Nederlandse taal versus Problemen met de Nederlandse taal;
5. WW 50+ jaar: werkzoekenden onder de 50 jaar met WW-uitkering versus overig;
6. WW 50-54 jaar: werkzoekenden vanaf 50 t/m 54 jaar met WW-uitkering versus overig;
7. WW 55+ jaar: werkzoekenden van 55 jaar en ouder met WW-uitkering versus overig;
8. WIA 35-min-indicatie: werkzoekenden met een WIA 35-min-indicatie (alle leeftijden) versus overig;

Aangezien de waarden van Cronbach's alfa en Guttman's lambda2 nauwelijks van elkaar verschillen, kijken we voor de subgroepen alleen naar de waarden van Guttman's lambda2 (zie tabel 7.4) en naar de item-rest correlaties.

Tabel 7.4: Guttman's lambda2 voor de factoren, weergegeven per subgroep op basis van achtergrondkenmerken.

Factor	Geslacht		Opleiding		Nederlandse Nationaliteit		Problemen Nederlands	
	man	vrouw	Laag	mi/ho	Nee	ja	Geen	Wel
Visie op terugkeer naar werk	0,81	0,79	0,79	0,80	0,66	0,80	0,80	0,67
Werkzoekgedrag m.b.t. direct contact werkgevers	0,73	0,69	0,71	0,72	0,70	0,72	0,72	0,68
Gezondheidsbeleving	0,85	0,86	0,86	0,85	0,82	0,86	0,85	0,84
Acceptatiebereidheid m.b.t. afwijkende tijden	0,90	0,83	0,86	0,87	0,80	0,87	0,87	0,83
Acceptatiebereidheid m.b.t. onwenselijk werk	0,70	0,71	0,68	0,72	0,61	0,71	0,71	0,63
Werkzoekintentie	0,70	0,80	0,79	0,80	0,83	0,80	0,80	0,80
Werkzoekattitude m.b.t. voordeel/aangenaamheid	0,77	0,78	0,79	0,77	0,73	0,77	0,77	0,78
Werkzoekattitude m.b.t. nut en noodzaak	0,81	0,80	0,82	0,80	0,74	0,81	0,78	0,81
Externe variabele attributie	0,70	0,70	0,72	0,70	0,72	0,70	0,70	0,71
Self-efficacy m.b.t. sollicitatievoorbereiding	0,75	0,77	0,71	0,76	0,77	0,76	0,76	0,71

<sup>31</sup> Een WW'er met een WIA 35-min-indicatie is afgewezen voor een WIA-uitkering (WIA: Wet werk en inkomen naar arbeidsvermogen). Deze mensen zijn eerst twee jaar ziek geweest, maar omdat hun inkomensverlies minder dan 35% is, hebben zij geen recht op een uitkering op grond van de WIA. Door de wijze van claimbeoordeling van de WIA kunnen sommige WW'ers met een WIA 35-min-indicatie toch ernstige beperkingen hebben die het vinden van een baan kunnen belemmeren. Werkzoekenden met een WIA 35-min-indicatie zijn ingedeeld in de klantgroep WIA 35-min ongeacht leeftijd (de WIA 35 min status is hierbij leidend).

Factor	WW tot 50 jaar		WW 50-54 jaar		WW55+		WIA35-min	
	nee	ja	nee	ja	nee	ja	nee	ja
Visie op terugkeer naar werk	0,83	0,74	0,80	0,79	0,76	0,83	0,76	0,83
Werkzoekgedrag m.b.t. direct contact werkgevers	0,73	0,70	0,72	0,73	0,71	0,72	0,71	0,72
Gezondheidsbeleving	0,87	0,84	0,86	0,85	0,85	0,86	0,85	0,86
Acceptatiebereidheid m.b.t. afwijkende tijden	0,87	0,87	0,87	0,86	0,87	0,87	0,87	0,87
Acceptatiebereidheid m.b.t. onwenselijk werk	0,72	0,69	0,70	0,70	0,70	0,74	0,70	0,74
Werkzoekintentie	0,79	0,81	0,81	0,79	0,81	0,77	0,81	0,77
Werkzoekattitude m.b.t. voordeel/aangenaam	0,77	0,77	0,77	0,78	0,77	0,76	0,77	0,76
Werkzoekattitude m.b.t. nut en noodzaak	0,84	0,77	0,80	0,80	0,78	0,85	0,78	0,85
Externe variabele attributie	0,67	0,70	0,70	0,68	0,70	0,67	0,70	0,67
Self-efficacy m.b.t. sollicitatievoorbereiding	0,75	0,76	0,76	0,76	0,76	0,73	0,76	0,73

De waarden van lambda2 liggen voor de meeste factoren tussen de 0,70 en 0,90 in de verschillende groepen. Geen van de waarden is lager dan 0,60. Een waarde tussen de 0,61 en 0,69 vinden we voor de volgende factoren:

- > Visie op terugkeer naar werk (subgroepen Niet-Nederlandse nationaliteit, Problemen met Nederlandse taal);
- > Werkzoekgedrag m.b.t. direct contact werkgevers (subgroepen vrouwen, Problemen met Nederlandse taal);
- > Acceptatiebereidheid m.b.t. onwenselijk werk (subgroepen Niet-Nederlandse nationaliteit, Problemen met Nederlandse taal, laagopgeleiden, WW tot 50 jaar) en
- > Externe variabele attributie (subgroepen WW tot 50 jaar, WW 50-54 jaar, WW 55+ jaar, WIA 35-min).

Binnen alle subgroepen zijn de item-rest correlaties hoger of gelijk aan 0,30 (zie bijlage I). Een item-rest correlatie van 0,30 of hoger kunnen we als goed beoordelen. Het betekent dat het betreffende item (min of meer) hetzelfde meet als de andere items behorend bij dezelfde factor (Evers, Lucassen, Meijers & Sijtsma, 2010).

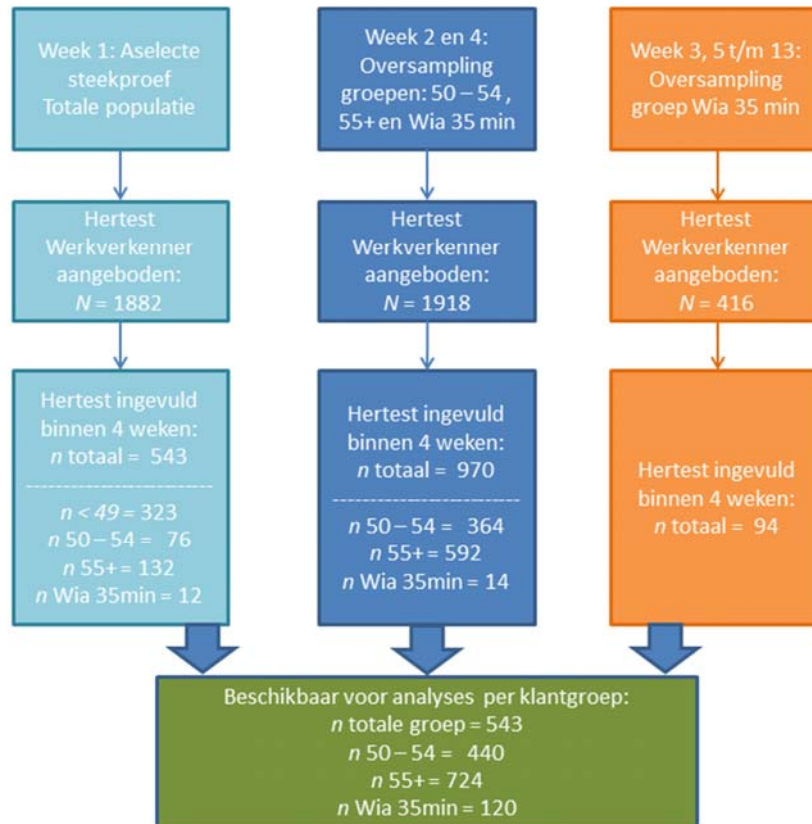
## 7.1.2 Test-hertest betrouwbaarheid

In deze paragraaf richten we ons letterlijk op de vraag of de test dezelfde scores oplevert bij herhaald meten (een hertest); dit wordt dan ook de test-hertest betrouwbaarheid genoemd. Een belangrijke aanname bij het bepalen van de test-hertest betrouwbaarheid is dat het gemeten concept niet verandert over tijd. Dit is aannemelijk voor diagnostische testen, zoals bijv. de Nederlandse Persoonlijkheds Vragenlijst (NPV; Luteijn et al., 2000). Bij diagnostische testen wordt verwacht dat de correlatie tussen de test en de hertest groter of gelijk is aan 0,70 (Evers, Lucassen, Meijer & Sijtsma, 2010). De zachte factoren van de Werkverkenner betreffen echter attitudes, motivatie en intentie op het gebied van het zoeken naar werk. Het ligt voor de hand om aan te nemen dat de scores op deze concepten enigszins fluctueren over tijd. We verwachten daarom voor deze factoren iets lagere test-hertest betrouwbaarheden (tussen de 0,60 en 0,70).

### 7.1.2.1 Beschrijving steekproef voor bepalen van test-hertest betrouwbaarheid

Voor het bepalen van de test-hertest betrouwbaarheid hebben we een aselechte steekproef getrokken uit de totale groep respondenten die de vragenlijst van de Werkverkenner hadden

ingevuld ( $N = 53.238$ , zie hoofdstuk 2). De uitnodiging om de Werkverkenner een tweede keer in te vullen volgde twee weken nadat de respondent de vragenlijst voor de eerste keer had ingevuld. De respondent kreeg twee weken de tijd om de vragenlijst voor een tweede keer in te vullen. De periode tussen het invullen van de twee vragenlijsten is dus minimaal twee en maximaal vier weken. Deze periode is daarmee lang genoeg om herinneringseffecten uit te schakelen, maar kort genoeg om het optreden van fluctuaties op bepaalde schalen te beperken. Zowel de eerste als tweede afname van de vragenlijst vinden plaats binnen de eerste 12 weken van werkloosheid.



Figuur 7.1: Stroomdiagram van de steekproeftrekking voor het test-hertest onderzoek.

In een periode van één week (31 januari t/m 7 februari, 2015) kreeg een aselecte groep respondenten de vragenlijst een tweede keer aangeboden in zijn/haar digitale werkmap. Dit waren in totaal 1882 werkzoekenden verspreid over de 11 koplopervestigingen van UWV die meededen (zie hoofdstuk 2). Van deze 1882 werkzoekenden hebben 543 respondenten de Werkverkenner voor een tweede keer ingevuld (respons = 29%).

Omdat we ook de test-hertest betrouwbaarheid wilden onderzoeken voor verschillende klantgroepen van UWV hebben we in de daaropvolgende weken (8 februari t/m 14 juni 2015) een gestratificeerde steekproef getrokken met als strata de specifieke klantgroepen van UWV (i.e. 50-54, 55+ en WIA 35-min). Een stroomdiagram met een verdeling over de verschillende groepen is weergegeven in figuur 7.1. We streefden naar minimaal 150 respondenten per klantgroep. Voor de klantgroepen 50-54 en 55+ hadden we in 3 weken genoeg respondenten. Omdat de groep WIA 35-min vrij klein was, hebben we voor deze

groep extra weken de tijd genomen; het aantal van 150 respondenten werd voor deze groep uiteindelijk echter niet gehaald.

Van de in totaal 4216 WW'ers die de vragenlijst een tweede keer aangeboden kregen, hebben 1620 WW'ers deze ingevuld binnen twee tot vier weken nadat zij de vragenlijst de eerste keer invulden. Dit komt neer op een responspercentage van 38,4%. Hiervan behoren 543 WW'ers tot de aselect getrokken steekproef, 440 tot de klantgroep werkzoekenden van 50 - 54 jaar, 724 tot de groep werkzoekenden van 55 jaar en ouder en 120 werkzoekenden tot de WIA 35-min groep.

#### 7.1.2.2 Representativiteit van de steekproef

**Conclusie:** De respons van de aselect getrokken steekproef van het test-hertest bestand ( $n = 543$ ) is representatief voor de totale groep respondenten die de Werkverkenner de eerste keer hadden ingevuld ( $N = 53.238$ ). In de rest van dit rapport verwijzen we dan ook naar deze groep als de "WW-representatieve groep".

We hebben getoetst of de respons van de aselect getrokken steekproef (i.e. de netto steekproef) van het test-hertest onderzoek ( $n = 543$ ) representatief was voor de populatie. Hiervoor hebben we de achtergrondkenmerken van de netto steekproef vergeleken met de achtergrondkenmerken van de referentiegroep. Deze referentiegroep betrof de totale groep respondenten in het onderzoekbestand die de Werkverkenner de eerste keer hadden ingevuld (i.e. de referentiegroep;  $N = 53.238$ );

Om de representativiteit te bepalen hebben we voor de categorische kenmerken getoetst of de geobserveerde verdeling in de netto steekproef gelijk is aan de verwachte verdeling (i.e. de waarden van de referentiegroep); dit deden we met een non-parametrische toets, waarbij de percentages van de referentie groep werden gebruikt om de verwachte aantallen (i.e. expected frequencies) te berekenen in de netto steekproef. Voor de continue kenmerken hebben we een "one sample  $t$ -test" uitgevoerd met gebruikmaking van het gemiddelde van de referentiegroep als verwachte waarde.

De analyses laten zien dat er geen verschillen zijn tussen de test-hertest netto steekproef en de referentiegroep in de verdeling van status werkhervatting, WIA 35-min-indicatie en Nationaliteit (tabel 7.5). Voor leeftijd (gemiddelde en leeftijdsgroepen), geslacht en opleiding vinden we wel significante verschillen. In termen van effectgrootte zijn deze verschillen klein (Cramer's  $V < 0,30$  en Cohen's  $d < 0,30$ ; zie par. 2.3 voor meer toelichting over de effectgroottes). De gemiddelde leeftijd ligt in het test-hertestbestand iets hoger dan de referentiewaarde. Dit wordt veroorzaakt door het relatief grotere aantal 55-plussers (25,6%) dat mee heeft gedaan aan het test-hertest onderzoek.

We concluderen dat de netto steekproef van het test-hertest onderzoek als representatief kan worden gezien voor het totale onderzoekbestand dat gebruikt is. In de rest van dit rapport verwijzen we dan ook naar deze groep als de "WW-representatieve groep".

Tabel 7.5: Resultaten van de representativiteitsanalyse van het test-hertest onderzoek. Vergelijking van de verdelingen op de achtergrondkenmerken tussen de netto steekproef van het test-hertest bestand (i.e. de respons van de aselechte steekproef) en de totale groep respondenten die de Werkverkenner de eerste keer hadden ingevuld in het onderzoeksbestand (referentiegroep).

Variabele	Test-hertest: (n = 543)	Referentie-groep (N = 53.238)	Toetsgegevens	Effectgroottes (Cramer's V of Cohen's d)
Werkhervatting binnen 12 mnd (%)			$\chi^2(1) = 0,13; p = 0,72^1$	V = 0,02
Geen werk hervat binnen 12 maanden	47,0	47,8		
Werk hervat binnen 12 maanden	49,7	52,2		
Onbekend <sup>1</sup>	3,3	-		
Leeftijd (M ± SD)	44,9 ± 11,2	42,1 (11,6)	$t(542) = 5,7; p < 0,01$	d = 0,24
Leeftijd (%)			$\chi^2(1) = 29,76; p < 0,01$	V = 0,23
t/m 27 jaar	8,3	13,0		
28-49 jaar	51,7	56,2		
50-54 jaar	14,4	12,7		
≥ 55 jaar	25,6	18,0		
Geslacht (% vrouw)	44,9	50,9	$\chi^2(1) = 7,7; p < 0,01$ ,	V = 0,12
Opleiding (%)			$\chi^2(3) = 10,87; p < 0,01^1$	V = 0,14
Laag	24,9	20,0		
Midden	50,3	51,3		
Hoog	23,8	28,5		
Onbekend <sup>1</sup>	1,1	0,2		
WIA 35-min (%)	2,7	2,0	$\chi^2(1) = 0,12; p = 0,73$	V = 0,02
Nationaliteit (%)			$\chi^2(3) = 1,10; p = 0,78^1$	V = 0,05
Nederlandse 1e nationaliteit en geen 2e nationaliteit of een westerse 2e nationaliteit (muv Pools)	91,9	95,6		
Poolse 1e of 2e nationaliteit	1,5	1,3		
Niet-westerse 1e of 2e nationaliteit (grotendeels Turks of Marokkaans)	1,5	1,1		
Westerse 1e nationaliteit anders dan Nederlandse	1,7	1,8		
Onbekend <sup>1</sup>	3,5	0,1		

M = gemiddelde, SD = standaard deviatie, dikgedrukt duidt op een significant toetsresultaat ( $p < 0,05$ ).

<sup>1</sup>. De categorie 'onbekend' is niet meegenomen in de toets;

### 7.1.2.3 Test-hertest betrouwbaarheid van de factoren van de Werkverkenner

**Conclusie:** Op basis van de waarden van de test-hertest correlaties concluderen we dat de test-hertest betrouwbaarheid van de meeste factoren voldoende is voor de totale WW-representatieve groep, de normgroep en de subgroepen. De test-hertest correlaties van de meeste factoren in Werkverkenner 2.0 zijn ruim voldoende ( $r \geq 0,70$ ).

We hebben zowel voor de WW-representatieve groep als voor de groep werkhervatters (de normgroep) en niet-werkhervatters gekeken naar de test-hertest betrouwbaarheid. Tevens hebben we voor de drie specifieke klantgroepen van UWV de test-hertest betrouwbaarheidsmaten berekend. In tabel 7.6 staat een overzicht van de test-hertest



betrouwbaarheid voor al deze groepen. We vinden voor de meeste factoren een voldoende samenhang tussen de twee metingen ( $r \geq 0,60$ ) voor alle groepen. Voor de meeste factoren van Werkverkenner 2.0 die gebruikt worden in de voorspelregel vinden we voor alle groepen een ruim voldoende test-hertest betrouwbaarheid ( $r \geq 0,70$ ), met uitzondering van Balans voor- en nadelen van niet werken.

Een tweetal factoren laten een onvoldoende betrouwbaarheid ( $r < 0,60$ ) in meer dan twee groepen zien: de factor 'Werkzoekintentie' heeft een onvoldoende betrouwbaarheid voor alle groepen en de factor 'Belemmering: financiële problemen' voor de totale groep en vier subgroepen. Concreet betekent dit dat de individuele scores op deze factoren fluctueren (bijv. een hogere score op werkzoekintentie op de eerste test gaat niet samen met een hogere score op werkzoekintentie op de tweede test na twee tot en met vier weken). Dit kan tenminste twee mogelijke oorzaken hebben: ofwel de vraag is onduidelijk gesteld en de werkzoekende vult daardoor "zomaar" iets in ofwel het construct dat gemeten wordt fluctueert sterk over de tijd (bijv. intentie om werk te zoeken).

Uit tabel 7.6 blijkt tevens dat de groep WIA 35-min het vaakst factoren met een onvoldoende betrouwbaarheid laat zien (vijf in totaal). Omdat slechts één van deze vijf factoren in Werkverkenner 2.0 zit (i.e. Balans voor-en nadelen niet werken), heeft dit geen noemenswaardige gevolgen op het toekomstige gebruik van de Werkverkenner.

Tabel 7.6: Overzichtstabel van de test-hertest betrouwbaarheid ( $r$ ) voor de totale groep (i.e. de WW-representatieve groep), de groep werkhervatters (de normgroep), de groep niet-werkhervatters en de drie afzonderlijke klantgroepen. Groen gearceerde cellen wijzen op een Pearson  $r \geq 0,60$ , rood gearceerde op een Pearson  $r < 0,60$ .

WV 2.0	Factor	Totaal	Status werkhervatting binnen 12 maanden		Klantgroepen		
			Werk hervat	Werk niet hervat	50 - 54 jaar	55+ jaar	WIA 35-min
		n=543	n=270	n=255	n=440	n=724	n=120
<i>Factoren gemeten met meerdere items</i>							
ja	Visie op terugkeer naar werk	0,76	0,70	0,75	0,73	0,74	0,70
ja	Werkzoekgedrag m.b.t. direct contact werkgevers	0,78	0,73	0,80	0,76	0,78	0,78
ja	Gezondheidsbeleving	0,87	0,82	0,89	0,86	0,88	0,90
ja	Acceptatiebereidheid m.b.t. afwijkende tijden	0,84	0,83	0,84	0,84	0,88	0,85
nee	Acceptatiebereidheid m.b.t. onwenselijk werk	0,75	0,76	0,73	0,72	0,77	0,73
nee	Werkzoekintentie	0,48	0,43	0,49	0,47	0,56	0,42
nee	Werkzoekattitude m.b.t. nut en noodzaak	0,65	0,57	0,72	0,54	0,70	0,65
nee	Werkzoekattitude m.b.t. voordeel en aangenaamheid	0,70	0,66	0,73	0,69	0,70	0,74
nee	Externe variabele attributie	0,69	0,73	0,61	0,60	0,61	0,50
nee	Self-efficacy m.b.t. sollicitatievoorbereiding	0,60	0,66	0,55	0,60	0,62	0,52
<i>Factoren gemeten met 1 item</i>							
ja	Algemeen werkvermogen	0,79	0,64	0,86	0,77	0,82	0,80
ja	Aantal uren in staat te werken	0,91	0,92	0,90	0,84	0,89	0,83
ja	Balans voor- en nadelen van niet werken <sup>1</sup>	0,66	0,62	0,69	0,61	0,77	0,58
ja	Acceptatiebereidheid m.b.t. voltijd baan <sup>2</sup>	0,82	0,72	0,86	0,83	0,84	0,81
ja	Financiële noodzaak <sup>2</sup>	0,72	0,69	0,71	0,74	0,77	0,75
nee	Belemmering: financiële problemen <sup>2</sup>	0,53	0,41	0,60	0,44	0,53	0,54
nee	Belemmering: gebrek kinderopvang <sup>2</sup>	0,79	0,76	0,79	0,51	0,73	0,76
nee	Subjectieve norm familie en partner	0,68	0,61	0,74	0,66	0,72	0,72

<sup>1</sup> Voor deze factor, gemeten op een nominale schaal, is de phi coefficient berekend.

<sup>2</sup> Voor deze factoren, gemeten op een ordinale schaal, is Spearman's rho berekend.

### 7.1.2.4 Invloed van tijd op de test-herstest betrouwbaarheid

**Conclusie:** Het moment van invullen heeft geen invloed op de samenhang tussen de scores op beide meetmomenten. We hoeven in de analyses van de test-herstest daarom geen rekening te houden met de tijd tussen de twee meetmomenten van de Werkverkenner.

We hebben ook gekeken of de tijd tussen de metingen van invloed is op de samenhang. Hiervoor hebben we aparte correlaties berekend voor werkzoekenden in de WW-representatieve groep (dus niet voor de aparte groepen) die binnen twee weken na het invullen van de eerste meting aan de tweede meting hebben meegedaan ( $n = 414$ ) en voor WW-klanten die drie tot vier weken<sup>32</sup> tussen beide metingen hebben zitten ( $n = 130$ ). Met behulp van een Fisher  $r$ -to- $z$  transformatie<sup>33</sup> hebben we per factor gekeken of het verschil in correlaties significant is. De resultaten staan afgebeeld in tabel 7.7.

Tabel 7.7: Correlaties tussen factoren op beide metingen, voor werkzoekenden uit de WW-representatieve groep ( $n = 543$ ) die binnen 2 respectievelijk 3-4 weken na het invullen van de eerste meting ook de tweede meting hebben ingevuld.

WV 2.0	Factor	Correlaties		Significant verschil in correlatie <sup>1</sup>
		2 weken tussen metingen ( $n = 413$ )	3-4 weken tussen metingen ( $n = 130$ )	
ja	Visie op terugkeer naar werk	,77	,73	$z = 0,90, p > 0,05$
ja	Werkzoekgedrag m.b.t. direct contact werkgevers	,78	,80	$z = -0,52, p > 0,05$
ja	Gezondheidsbeleving	,89	,83	$z = 2,30, p < 0,05$
ja	Acceptatiebereidheid m.b.t. afwijkende tijden	,85	,82	$z = 0,98, p > 0,05$
nee	Acceptatiebereidheid m.b.t. onwenselijk werk	,75	,74	$z = 0,22, p > 0,05$
nee	Werkzoekintentie	,48	,47	$z = 0,13, p > 0,05$
nee	Werkzoekattitude m.b.t. nut en noodzaak	,66	,63	$z = 0,51, p > 0,05$
nee	Werkzoekattitude m.b.t. voordeel en aangenaamheid	,72	,64	$z = 1,47, p > 0,05$
nee	Externe variabele attributie	,71	,63	$z = 1,44, p > 0,05$
nee	Self-efficacy (sollicitatievoorbereiding)	,63	,50	$z = 1,89, p > 0,05$

1. Fisher's  $r$ -naar- $z$  transformatie (tweezijdige toets)

<sup>32</sup> Omdat er slechts 19 werkzoekenden zijn die vier weken tussen beide metingen heeft, hebben we deze respondenten samengevoegd met de groep werkzoekenden die drie weken tussen de metingen heeft.

<sup>33</sup> Om te kunnen toetsen of twee correlaties significant van elkaar verschillen, worden de correlaties eerst getransformeerd. Hiervoor gebruiken we Fisher's- $r$ -naar- $z$  transformatie. Elke correlatie wordt eerst getransformeerd naar een  $z$ -waarde, waarna vervolgens deze  $z$ -waarden worden gebruikt om te toetsen of de correlaties verschillend zijn in de twee groepen. De  $r$ -naar- $z$  transformatie heeft weinig effect op lage correlaties ( $r$  zal vrijwel gelijk zijn aan  $z$ ), maar voor hogere correlaties zal de  $z$ -waarde groter zijn dan  $r$ . Dit heeft als gevolg dat een verschil tussen twee correlaties niet significant kan zijn bij lagere correlaties, terwijl een eenzelfde verschil een significant resultaat geeft voor hogere correlaties. (Dit zien we bijvoorbeeld terug bij de factor Gezondheidsbeleving (kleiner verschil tussen de correlaties, wel significant) versus Self-efficacy (groter verschil, niet significant)).

Zoals is te zien zijn alle correlaties op de factoren tussen beide metingen hoger in de groep waarbij er twee weken tussen beide metingen zit. Een uitzondering vinden we voor de factor 'Direct contact werkgevers': hier is de correlatie tussen de metingen iets hoger voor werkzoekenden met drie tot vier weken tussen de metingen. De verschillen in (absolute) correlaties zijn echter klein (tussen de 0,01 en 0,13) en niet significant. Voor de factor Gezondheidsbeleving zijn de correlaties tussen de meetmomenten wel significant verschillend voor werkzoekenden die twee en drie tot vier weken tussen de metingen hebben zitten. Echter, de correlaties zijn in beide groepen hoog. We vinden weliswaar een significant hogere correlatie op Gezondheidsbeleving voor de werkzoekenden die twee weken tussen de metingen hebben zitten ( $r = 0,89$ ) in vergelijking met werkzoekenden die drie tot vier weken tussen de metingen hebben zitten ( $r = 0,83$ ), maar de hoge correlaties geven aan dat de scores over tijd in *beide* groepen zeer stabiel zijn. We maken ons daarom geen zorgen over de significante bevinding voor deze factor. Aan de hand van deze extra analyse concluderen we daarom dat we in de analyses voor dit rapport geen rekening hoeven te houden met het moment van invullen van de hertest.

#### 7.1.2.5 Stabiliteit van de scores

We hebben tevens gekeken naar de stabiliteit van de scores over tijd (op groepsniveau, dus gemiddeld genomen). Voor de WW-representatieve groep geldt dat de gemiddelde scores op de twee tijdstippen nauwelijks van elkaar verschillen (tabel 7.8). Dit is gunstig, omdat we hiermee hebben aangetoond dat over een tijdsaspect van twee tot vier weken de gemiddelde scores op de Werkverkenner stabiel zijn.

Tabel 7.8: Gemiddelden en effectgroottes tussen de twee metingen voor de WW-representatieve groep (n = 543).

Factor	Meting 1		Meting 2		Effect-grootte <sup>1</sup>
	Gemiddelde	SD	Gemiddelde	SD	
Self-efficacy m.b.t. sollicitatievoorbereiding	3,9	0,6	3,9	0,6	0,09
Werkzoekattitude m.b.t. voordeel en aangenaamheid	3,3	0,8	3,3	0,8	0,02
Werkzoekattitude m.b.t. nut en noodzaak	4,3	0,7	4,3	0,7	0,06
Visie op terugkeer naar werk	3,5	0,8	3,5	0,8	0,04
Werkzoekintentie	4,1	0,7	4,0	0,7	0,11
Gezondheidsbeleving	4,0	0,7	4,0	0,8	0,00
Externe variabele attributie	3,3	0,8	3,4	0,8	0,05
Direct contact werkgevers	1,9	0,6	1,9	0,6	0,03
Acceptatiebereidheid m.b.t. afwijkende tijden	2,7	1,3	2,7	1,3	0,08
Acceptatiebereidheid m.b.t. onwenselijk werk	2,9	0,9	3,0	0,9	0,15

1. Cohens'd: Effectgrootte tussen 0 en |0,20| wijst op een (zeer) klein (verwaarloosbaar) effect, rond de |0,50| op een middelgroot effect en waarden boven de |0,80| duiden op een groot effect (Cohen, 1988). SD = standaarddeviatie.

## 7.2 Construct validiteit

**Algehele conclusie:** De constructvaliditeit van de tien factoren in de Werkverkenner die met meerdere items worden gemeten is goed. De factoren zijn bruikbaar voor de gehele onderzoekspopulatie. Dit blijkt uit confirmatieve factor analyses.

**Conclusie confirmatieve factor analyse:** Uit de confirmatieve factoranalyse voor de totale groep blijkt dat de veronderstelde factorstructuur past goed bij de geobserveerde data. Voor alle items werden factorladingen van voldoende grootte gevonden.

**Conclusie invariantie factorstructuur en mogelijke item bias:** De veronderstelde factorstructuur is houdbaar voor groepen onderscheiden naar geslacht, opleidingsniveau, nationaliteit, problemen Nederlandse taal, leeftijd en WIA 35-min. Verder zijn er geen aanwijzingen voor item bias. Aangenomen mag worden dat de factoren hetzelfde worden geïnterpreteerd in de onderscheiden groepen en dat we de scores op de factoren kunnen vergelijken tussen de groepen.

**Conclusie soortgenotentest:** De veronderstelde samenhang (i.e. convergente validiteit) tussen schalen van de Werkverkenner en de Arbeidsmarktmotivatie en Werkzoekgedrag Vragenlijst (AWV), vinden we voor een aantal schalen terug in de data. Voor de meeste schalen vinden we echter een minder hoge samenhang dan we van tevoren hadden verwacht en ook niet altijd met de verwachte AWV-schaal. De AWV sluit niet op alle gebieden aan bij de Werkverkenner. Daarnaast zijn de Werkverkenner en de AWV niet op dezelfde dag ingevuld, wat een invloed kan hebben gehad op de samenhang tussen de schalen.

Bij het aantonen van de validiteit van een test gaat het erom of de test meet wat hij zou moeten meten. Er zijn een aantal vormen van validiteit. In deze paragraaf beschrijven we de constructvaliditeit, ook wel begripsvaliditeit genoemd. Een manier om de constructvaliditeit van een test te onderzoeken is om te bekijken of de vooraf veronderstelde structuur van een test (i.e. de zogenoemde hypothetische structuur) opgaat voor de geobserveerde gegevens. Dit zullen we doen met behulp van confirmatieve factor analyse. Een tweede manier (nauw samenhangend met de eerste manier) is het onderzoeken van de invariantie van de factorstructuur en mogelijke itembias bij verschillende subgroepen. Een derde manier om constructvaliditeit aan te tonen is door te bepalen in hoeverre de scores op de constructen (i.e. factoren) van de Werkverkenner samenhangen met gelijksoortige constructen gemeten door een andere test (i.e. een soortgenotentest). Deze derde vorm van constructvaliditeit zal ook aan bod komen.

### 7.2.1 *Confirmatieve factoranalyse voor de totale groep*

De zachte factoren weergegeven in tabel 7.1 zijn gemeten met 36 items. Aangenomen wordt dat de geobserveerde items dienen als empirische indicatoren voor deze hypothetische factoren (andere benamingen hiervoor zijn: achterliggende factoren of latente constructen). Om te bekijken of de verwachte factorstructuur ook terug te vinden is in de geobserveerde data, wordt het hypothetische model getoetst met een confirmatieve factoranalyse. Factoren die slechts door één item worden gemeten kunnen hierbij niet worden meegenomen. Dit zijn namelijk feitelijk geen achterliggende factoren, omdat het hierbij niet gaat om een groep van items die samen hetzelfde construct beogen te meten. Deze items werden wel opgenomen in de analyses voor de voorspelling van werkhervatting en werkloosheidsduur.

We hebben in het huidige onderzoek getoetst of de afzonderlijke gezondheidsitems 'psychisch werkvermogen', 'lichamelijk werkvermogen', 'ervaren gezondheid' en 'te ziek

voelen om te werken' alle over hetzelfde construct 'gezondheidsbeleving' gaan en hierdoor kunnen worden samengevoegd tot 1 factor. Het item 'Hoeveel punten zou u geven aan uw werkvermogen?' zou theoretisch ook bij deze factor passen, maar nemen we door de afwijkende antwoordschaal (rapportcijfer van 1 t/m 10) niet op in het factormodel. Uiteindelijk zullen namelijk de scores op de items die samen een factor vormen worden samengenomen tot een totaal- of schaalscore. Het construeren van deze zogenoemde schaalscore wordt afgeraden voor items met verschillende antwoordschalen.

Als we de zachte factoren in tabel 7.1 die worden gemeten door één item verwijderen, komen we in totaal uit op 10 achterliggende factoren die geschikt zijn voor een confirmatieve factor analyse.

In tegenstelling tot de veelgebruikte exploratieve factor analyse is het bij confirmatieve factor analyse niet mogelijk dat items een lading hebben op meerdere achterliggende factoren. Bij confirmatieve factor analyse laadt ieder item slechts op de factor waar het item volgens de hypothetische structuur bij hoort. De vraag die aan de analyses ten grondslag lag was dan ook: "Passen de waargenomen correlaties tussen de items bij de hypothetische factorstructuur bestaande uit 10 factoren?". Van te voren is gespecificeerd welke items bij welke factoren horen. In de analyses mochten de factoren onderling met elkaar samenhangen.

#### 7.2.1.1 *Fit van een factormodel*

Of het factormodel past bij de geobserveerde gegevens wordt weergegeven door fit indices. Er zijn verschillende fit indices beschikbaar en het wordt aangeraden om verschillende indices te gebruiken om de uiteindelijke fit van het model te bepalen. Een chi-kwadraattoets toetst de discrepantie tussen het veronderstelde model en de geobserveerde gegevens. Een niet-significante toets geeft aan dat het veronderstelde model en het geobserveerde model niet (veel) van elkaar verschillen, wat een goed verondersteld model inhoudt. Een nadeel van deze toets is dat de uitkomst onder andere afhangt van de grootte van de steekproef. Bij een grote steekproef (zoals hier het geval is) zullen kleine afwijkingen van het factormodel al statistisch significant zijn en zullen ook vrij goede modellen worden verworpen. We kijken daarom ook naar goodness-of-fit indexen. Deze geven ook de fit van een model aan, maar steekproefgrootte speelt bij deze maten geen rol. Voorbeelden zijn de Comparative Fit Index (CFI), de Normed Fit Index (NFI) en de Tucker-Lewis Index, waarbij geldt dat waarden boven 0,90 en  $\geq 0,95$  duiden op respectievelijk een acceptabele en goede fit. Een waarde kleiner dan 0,05 voor de root mean square error of approximation (RMSEA) duidt op een goede fit (Hu & Bentler, 2009). Van RMSEA en CFI is bekend dat ze rekening houden met spaarzaamheid (Bentler, 1990; Kline, 2004).

#### 7.2.1.2 *Resultaten van de confirmatieve factor analyse voor totale groep: fit indices*

**Conclusie:** Op basis van de CFI ( $\geq 0,95$ ), NFI ( $\geq 0,95$ ), TLI ( $\geq 0,95$ ) en RMSEA ( $< 0,05$ ) concluderen we dat er sprake is van een goede fit. Dit betekent dat het model met tien factoren goed past bij de geobserveerde data. Het voorstel om de vier afzonderlijke factoren 'psychisch werkvermogen', 'lichamelijk werkvermogen', 'ervaren gezondheid' en 'te ziek voelen om te werken' samen te nemen tot een factor 'gezondheidsbeleving' blijkt goed houdbaar.

De fit indices zijn weergegeven in tabel 7.9 en laten zien dat het model met tien factoren een goed gespecificeerd model is.

Tabel 7.9: Model fit indices van de confirmatieve factor analyse met tien factoren<sup>a</sup>.

	<i>chi-kwadraat</i>	df	CFI	NFI	TLI	RMSEA
Factormodel	21476,32	279	0,96	0,96	0,95	0,04

<sup>a</sup>. De chi-kwadraat toets is significant, maar deze is al gauw significant bij een grote steekproef. Daarom wordt deze hier buiten beschouwing gelaten in de evaluatie van het factormodel. df = degrees of freedom (vrijheidsgraden), CFI = comparative fit index, NFI = normed fit index, TLI = Tucker-Lewis index, RMSEA = Root mean square error of approximation.

In tabel 7.9 zijn globale fitmaten gepresenteerd. Daarnaast is het ook mogelijk om gedetailleerder naar de fit van het model op itemniveau te kijken. Daarvoor hebben we naar de correlatiematrix gekeken en meer specifiek naar de residuen. Een residu is in dit geval het verschil tussen de geobserveerde correlatie en de teruggeschatte correlatie door het model. Hoe dichterbij 0 het residu ligt, hoe beter het model erin slaagt om de correlatie terug te schatten. Wanneer de fit van het model niet acceptabel zou zijn, dan kunnen residuen inzicht geven waar de verschillen liggen tussen de geobserveerde en de geschatte correlaties en waar het hypothetisch model dus mogelijk verbeterd kan worden. Omdat de algehele fit van het model goed is, hebben de waarden van de residuen voor de huidige analyses geen verdere implicaties. In ons geval waren alle residuen  $\leq |0,13|$ , wat impliceert dat de geobserveerde correlaties niet meer dan 0,13 afwijken van de geschatte correlaties door het model. Het hoogste residu vonden we tussen de items 'Ik weet hoe ik de komende maand kan proberen werk te vinden' (factor Werkzoekintentie) en 'Ik verwacht snel zelf een baan te vinden' (factor Visie op terugkeer naar werk) (residu = 0,13). Dit betekent dat de geobserveerde correlatie tussen deze twee items iets hoger ligt dan de correlatie die door het model wordt geschat.

**Aanbevelingen:** In het huidige voorgestelde factormodel bestaan een aantal factoren slechts uit twee items, zoals bijvoorbeeld de factor Acceptatiebereidheid m.b.t. onwenselijk werk. Voor toekomstig onderzoek bevelen we aan om meer dan twee items per factor in het model op te nemen. Een minimum van drie items per factor wordt namelijk aangeraden voor een stabiele factor (MacCallum, 1995). Daarnaast heeft het de voorkeur om ervoor te zorgen dat items behorende bij een factor hetzelfde aantal antwoordcategorieën hebben. Voor het construeren van schaalscores is dit van belang, omdat alle items dan dezelfde bijdrage (i.e. gewicht) aan de factorscore leveren. In de huidige analyses is bijvoorbeeld Algemeen werkvermogen niet opgenomen in de confirmatieve factor analyse, omdat de antwoordcategorieën van dit item (een schaal van 1-10) verschillen van de items die bij de factor 'gezondheidsbeleving' passen (een schaal van 1-5). In vervolgonderzoek kan hier bij het opstellen van een vragenlijst al rekening mee worden gehouden door items met hetzelfde aantal antwoordschalen te selecteren, zodat er geen items buiten de factoranalyse vallen.

### 7.2.1.3 Resultaten van de confirmatieve factor analyse voor totale groep: factorladingen

**Conclusie:** Voor alle factoren en de bijbehorende items werden ladingen van voldoende grootte gevonden (factorladingen tussen  $|0,53|$  en  $|0,90|$ ). Dit betekent dat we de items behorende bij een factor samen kunnen nemen, waarmee we één score krijgen die een goede samenvattende weergave van de betreffende items is.

Nu blijkt dat het theoretisch model goed bij de data past, kijken we naar de geschatte (gestandaardiseerde) factorladingen. Factorladingen geven het verband (oftewel, de correlatie) aan tussen het geobserveerde item en de factor. We kijken hier naar de gestandaardiseerde ladingen, omdat deze ladingen makkelijk te interpreteren zijn. Gestandaardiseerde ladingen kunnen waarden aannemen tussen de -1 en 1. Een

factorlading van |1| duidt op een perfecte correlatie, terwijl een lading van 0 aangeeft dat er geen verband is tussen een item en de factor. Een factorlading dichtbij 1 geeft aan dat er een sterk verband is tussen de score op het geobserveerde item en de factorscore en dat is een gewenste uitkomst. Een waarde dichtbij 0 geeft aan dat de factor geen goede representatie is van het item (m.a.w. dat het item niet goed past bij de factor). Negatieve ladingen geven een tegengesteld verband aan: een hoge (of lage) score op een item geeft een lage (of hoge) score op de factor. Een factorlading van  $\geq |0,50|$  wordt over het algemeen als ondergrens genomen (Tabachnick & Fidell, 2001). De factorladingen zijn weergegeven in tabel 7.10. Alle factorladingen zijn voldoende groot (dat wil zeggen,  $> |0,50|$ ).

Tabel 7.10: Factoren met de bijbehorende items en hun factorladingen.

Factor	Item	Factorlading, gestandaardiseerd
Visie op terugkeer naar werk	Ik denk dat het lang zal duren voordat ik werk vind (omgepooled)	0,83
	Ik ben vaak afgewezen, ik verwacht geen werk meer te vinden (omgepooled)	0,72
	Ik verwacht snel zelf een baan te vinden	0,71
Werkzoekgedrag m.b.t. direct contact werkgevers	Ik heb sollicitatiegesprekken gevoerd	0,67
	Ik heb met werkgevers gebeld	0,67
	Ik heb werkgevers bezocht	0,70
Gezondheidsbeleving	Hoe beoordeelt u uw werkvermogen op dit moment, als u kijkt naar de psychische (geestelijke) eisen die werk aan u zou stellen?	0,77
	Hoe beoordeelt u uw werkvermogen op dit moment, als u kijkt naar de lichamelijke eisen die werk aan u zou stellen?	0,86
	Wat vindt u, over het algemeen, van uw gezondheid?	0,81
	Ik voel me te ziek om te werken (omgepooled)	0,67
Acceptatiebereidheid m.b.t. afwijkende tijden	In hoeverre wilt u werken op onregelmatige werktijden (werken op verschillende tijden)?	0,85
	In hoeverre wilt u buiten kantoor tijden werken ('s avonds en 's nachts)?	0,90
Acceptatiebereidheid m.b.t. onwenselijk werk	Wilt u werk onder uw niveau?	0,76
	Wilt u werk voor een lager salaris?	0,71
Self-efficacy (sollicitatie voorbereiding)	Ik kan goed informatie vinden over vacatures	0,67
	Ik kan mijn sterke en zwakte punten voor een baan uitleggen	0,74
	Ik kan goed een (digitale) sollicitatie schrijven	0,75
Werkzoekintentie	Ik ben van plan om er de komende maand alles aan te doen om werk te vinden	0,77
	Ik weet hoe ik de komende maand kan proberen werk te vinden	0,66
	Ik ben zeer gemotiveerd om komende maand werk te zoeken	0,87
Werkzoekattitude m.b.t. voordeel en aangenaamheid	Wat vindt u van het zoeken naar een betaalde baan, prettig of onprettig?	0,74
	Wat vindt u van het zoeken naar een betaalde baan,	0,85



Factor	Item	Factorlading, gestandaardiseerd
	nadelig of voordelig?	
Werkzoekattitude m.b.t. nut en noodzaak	Wat vindt u van het zoeken naar een betaalde baan, onnodig of nodig?	0,76
	Wat vindt u van het zoeken naar een betaalde baan, nutteloos of nuttig?	0,88
Externe variabele attributie	Ik denk dat je voor het krijgen van een baan vooral geluk moet hebben	0,77
	Ik denk dat je voor het krijgen van een baan vooral op het juiste moment op de juiste plaats moet zijn	0,67
	Ik denk dat de meeste mensen een baan vinden door de invloed van anderen	0,53

### 7.2.2 *Invariantie factor structuur en mogelijke item bias bij verschillende subgroepen*

In de voorgaande paragraaf hebben we geconcludeerd dat het model met tien factoren goed past bij de geobserveerde data van de totale groep. Nu willen we onderzoeken of dit ook het geval is voor verschillende groepen. Dit doen we door het factormodel simultaan (confirmatief) te schatten in twee groepen. Wanneer dezelfde factorstructuur past in de twee groepen, dan is het aannemelijk dat de factoren hetzelfde worden geïnterpreteerd in de groepen. We spreken dan van *invariantie* van de factor structuur. Voor het aantonen van deze invariantie stellen we als eis dat de factorstructuur gelijk moet zijn in de twee groepen (d.w.z. dezelfde items laden op dezelfde 10 factoren), de factorladingen mogen echter wel variëren tussen de twee groepen. Als blijkt dat de onderliggende factorstructuur niet gelijk is, dan zijn de factoren niet zonder meer bruikbaar voor de algehele onderzoekspopulatie.

Tevens onderzoeken we in deze paragraaf de item bias. Er is sprake van item bias als een item een verschillende betekenis heeft voor verschillende groepen. De kans op een bepaald antwoord is dan mede afhankelijk van de groep waartoe iemand behoort. Een manier om item bias te onderzoeken is om het factormodel simultaan te schatten in twee of meer groepen, waarbij tevens de aanname wordt gedaan van gelijke factorladingen en gelijke intercepten voor de groepen (zie bijvoorbeeld Cheung & Rensvold, 2002). De aanname van gelijke factorladingen (ook wel zwakke invariantie genoemd) houdt in dat de factor dezelfde betekenis heeft in verschillende groepen. De aanname van gelijke intercepten (ook wel sterke invariantie genoemd) houdt in dat individuen die dezelfde score hebben op een factor, ook een vergelijkbare score op het bijbehorende item moeten hebben, ongeacht de groep waar ze in zitten (bijvoorbeeld man of vrouw). Het verschil in chi-kwadraat tussen bijvoorbeeld het model met gelijke factor ladingen en het model met gelijke factorladingen en gelijke intercepten kan worden getest met een likelihood ratio test. Een niet-significante toets geeft aan dat het meer restrictieve model en het minder restrictieve model niet (veel) van elkaar verschillen, wat in dit geval betekent dat de aanname van zowel gelijke factorladingen als gelijke intercepten in de beide groepen gerechtvaardigd is. Een nadeel van deze toets is echter (wederom) dat de uitkomst, onder andere, afhangt van de grootte van de steekproef. Bij een grote steekproef (zoals hier het geval is) zullen relatief kleine verschillen al statistisch significant zijn en zal dus de aanname van bijv. sterke invariantie snel verworpen worden. We kijken daarom wederom hoofdzakelijk naar de fitmaten (CFI, NFI en RMSE). Wanneer de confirmatieve factormodellen met de aanname van sterke invariantie een goede fit laat zien, mogen we aannemen dat er geen sprake is van item bias (He & Van den Vijver, 2012).

We concluderen dan dat de groepen dezelfde betekenis (i.e. interpretatie) geven aan de

factoren en aan de items die bij de factoren horen. Hierdoor kunnen we groepen vergelijken op hun factorscores. Bij een slechte fit zullen we de items die deze slechte fit veroorzaken nader onderzoeken.

We testen de invariantie van de factorstructuur en de item bias in de volgende groepen:

- > Geslacht: mannen versus vrouwen;
- > Opleiding: laag versus middel/hoog opgeleid;
- > Nationaliteit: Nederlandse versus niet Nederlandse nationaliteit;
- > Problemen met de Nederlandse taal: geen problemen met de Nederlandse taal versus Problemen met de Nederlandse taal;
- > Klantgroep WW 50- jaar: werkzoekenden onder de 50 jaar met WW-uitkering versus overig;
- > Klantgroep WW 50-54 jaar: werkzoekenden tussen 50 en 54 jaar met WW-uitkering versus overig;
- > Klantgroep WW 55+ jaar: werkzoekenden van 55 jaar en ouder met WW-uitkering versus overig;
- > Klantgroep WIA 35-min: werkzoekenden met een WIA 35-min-indicering (alle leeftijden) versus overig.

#### 7.2.2.1 Resultaten invariantie factorstructuur

**Conclusie:** De factorstructuur is houdbaar voor groepen onderscheiden naar geslacht, opleidingsniveau, nationaliteit, problemen Nederlandse taal, leeftijd en WIA 35-min. Aangenomen mag worden dat de factoren hetzelfde worden geïnterpreteerd in deze onderscheiden groepen. Dit is een belangrijke uitkomst, omdat hiermee is aangetoond dat de factoren bruikbaar zijn voor de gehele onderzoekspopulatie.

De fit maten van het model met tien factoren zijn vrijwel gelijk in de verschillende subgroepen (zie tabel 7.11). De fit maten CFI, TLI en RMSEA geven aan dat het model goed past. Dit betekent dat we kunnen aannemen dat de factorstructuur gelijk is in de verschillende subgroepen, dus het 10-factoren model past bijv. zowel in de groep met Nederlandse nationaliteit als in de groep met een niet-Nederlandse nationaliteit. Dit wordt ook wel configurele invariantie genoemd.

Tabel 7.11: Model fit indices van de confirmatieve factor analyses met tien factoren voor de subgroepen.

	<i>chi-kwadraat</i>	<i>df</i>	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>
Geslacht (man vs. vrouw)	22141,02	558 <sup>a</sup>	0,96	0,95	0,04
Opleiding (laag vs. middel/hoog)	20353,76	558	0,96	0,95	0,04
Nationaliteit (Nederlands vs. niet-Nederl.)	21698,85	558	0,96	0,95	0,04
Problemen Nederlandse taal (geen vs. wel)	21182,50	558	0,96	0,95	0,04
Klantgroep WW50- (ja vs. nee)	22079,03	558	0,96	0,95	0,04
Klantgroep WW50-54 (ja vs. nee)	21856,78	558	0,96	0,95	0,04
Klantgroep WW55+ (ja vs. nee)	22131,56	558	0,96	0,95	0,04
Klantgroep WIA35-min (ja vs. nee)	21812,02	558	0,96	0,95	0,04

a. Dit is tweemaal het aantal vrijheidsgraden van het geschatte model in de totale groep, omdat het model voor twee groepen simultaan is geschat.

*df* = vrijheidsgraden (degrees of freedom), *CFI*= comparative fit index, *NFI*= normed fit index, *TLI*= Tucker-Lewis index, *RMSEA*=Root mean square error of approximation.

### 7.2.2.2 Resultaten item bias

**Conclusie:** Er zijn geen aanwijzingen voor item bias. Dit betekent dat we scores op de factoren kunnen vergelijken tussen groepen.

De fit van het model met tien factoren, waarbij gelijke factorladingen en gelijke intercepten worden verondersteld in de twee groepen (bijvoorbeeld mannen vs. vrouwen; of laag opgeleiden vs. middel/hoog opgeleiden), is voor de verschillende achtergrond-kenmerken weergegeven in tabel 7.12. Uit de tabel is af te lezen dat het factormodel een goede fit laat zien,  $CFI \geq 0,95$ ,  $TLI \geq 0,95$  en  $RMSE \leq 0,05$ , voor vrijwel alle achtergrondkenmerken. Er zijn twee uitzonderingen: voor de klantgroep WW 0 – 50 en de klantgroep 55plus is de TLI gelijk aan 0,94. Omdat de twee andere fitmaten (CFI en RMSEA) wel goed zijn in deze klantgroepen, concluderen dat er geen duidelijke aanwijzingen zijn voor item bias.

Tabel 7.12: Model fit indices van de confirmatieve factor analyses met 10 factoren voor de achtergrondkenmerken, met de assumpties gelijke factorladingen en gelijke intercepten in de twee groepen.

	chi-kwadraat	df	CFI	TLI	RMSEA
Geslacht (man vs. vrouw)	23785,85	592 <sup>a</sup>	0,96	0,95	0,04
Opleiding (laag vs. middel/hoog)	24192,37	592	0,95	0,95	0,04
Nationaliteit (Nederlands vs. niet-Nederl.)	22607,37	592	0,96	0,95	0,04
Problemen Nederlandse taal (geen vs. wel)	22746,47	592	0,96	0,95	0,04
Klantgroep WW50- (ja vs. nee)	24969,67	592	0,95	0,94	0,04
Klantgroep WW50-54 (ja vs. nee)	22324,79	592	0,96	0,95	0,04
Klantgroep WW55+ (ja vs. nee)	24371,46	592	0,95	0,94	0,04
Klantgroep WIA35-min (ja vs. nee)	22545,52	592	0,96	0,95	0,04

a. Dit is iets meer dan tweemaal het aantal vrijheidsgraden van het geschatte model in de totale groep, omdat de factorladingen en de intercepten gelijk zijn gesteld in de twee groepen.  
df = vrijheidsgraden (degrees of freedom), CFI= comparative fit index, NFI= normed fit index, TLI= Tucker-Lewis index, RMSEA=Root mean square error of approximation.

### 7.2.3 Soortgenoten onderzoek

**Algehele conclusie:** In het soortgenotenonderzoek van de Werkverkenner vinden we samenhang met de Arbeidsmarktmotivatie en Werkzoekgedrag Vragenlijst (AWV), een vragenlijst die ook arbeidsmotivatie en het daadwerkelijk zoekgedrag van de werkzoekende meet. De veronderstelde samenhang tussen schalen van de Werkverkenner en de AWV vinden we voor een aantal schalen terug in de data. Voor de meeste schalen vinden we echter een minder hoge samenhang dan we van tevoren hadden verwacht en ook niet altijd met de verwachte AWV-schaal. De AWV sluit niet op alle gebieden aan bij de Werkverkenner. Daarnaast zijn de Werkverkenner en de AWV niet op dezelfde dag ingevuld, wat een invloed kan hebben gehad op de samenhang tussen de schalen. Op basis van het soortgenoten onderzoek hebben is er geen reden om schalen van de Werkverkenner niet mee te nemen in de analyses voor de voorspelling van werkhervatting.

Voor het bepalen van de constructvaliditeit (ook wel begripsvaliditeit genoemd) van de Werkverkenner wordt onderzocht of de Werkverkenner de constructen meet, die het beoogt te meten. Een onderdeel van constructvaliditeit is de convergente validiteit, waarbij er wordt gekeken naar de samenhang tussen twee vragenlijsten die hetzelfde begrip beogen te meten. Dit wordt ook wel een 'soortgenoten'-onderzoek genoemd. Hoe hoger de samenhang tussen de twee testen, hoe beter de convergente validiteit.

Het specifieke doel van dit soortgenoten onderzoek is het bepalen van de convergente validiteit van de zachte schalen van de Werkverkenner. De test waarmee de Werkverkenner vergeleken gaat worden, noemen we de soortgenotentest. De COTAN<sup>34</sup> stelt geen andere specifieke eisen aan de soortgenotentest dan een betrouwbaarheid van  $\geq .60$  en voldoende onderbouwing voor de validiteit van het instrument. Bij het zoeken naar een goede soortgenoten test voor de Werkverkenner is in eerste instantie in de COTAN testdatabase gekeken of er een test beschikbaar was, die inhoudelijk aansluit bij de Werkverkenner én een voldoende/goede beoordeling heeft. Er bleken echter geen vragenlijsten te zijn die inhoudelijk voldoende aansloten bij de verschillende aspecten van de Werkverkenner en minimaal een voldoende beoordeling hadden.

Aan de basis van de Werkverkenner liggen drie theoretische modellen ten grondslag over harde en zachte persoons- en situatietekenen en de relatie met werkhervatting (Brouwer et al., 2011). Dit zijn het multidisciplinaire model van Wanberg (Wanberg, Song & Hough, 2002), de theorie van gepland gedrag (TPB; Ajzen, 1991) en het motivatiemodel van Vroom (Vroom, 1964). Bij de formulering van de vragen voor de zachte kenmerken van de Werkverkenner zijn verschillende, onder andere Engelstalige, vragenlijsten als uitgangspunt genomen (o.a. Blau, 1994; Kopelman, Rovenpor & Millsap, 1992; Furnham & Rawles, 1996). Vanuit het perspectief van de theorie van gepland gedrag heeft in Nederland Van Hooft onderzoek gedaan naar verschillende vormen van werkzoekgedrag, o.a. in de context van re-integratie (Van Hooft, 2004; Van Hooft, Born, Taris, Van der Flier, 2004). De vragenlijsten die als uitgangspunt voor de ontwikkeling van de Werkverkenner zijn gebruikt, zijn veelal dezelfde als de vragenlijsten in het onderzoek van Van Hooft (2004). De modellen en vragenlijsten uit het onderzoek van Van Hooft (2004) vormen tevens de basis voor de Arbeidsmarktmotivatie en Werkzoekgedrag Vragenlijst (AWV), ontwikkeld door NOA. Deze vragenlijst sluit daarom inhoudelijk op meerdere punten aan bij de Werkverkenner. De AWV meet de arbeidsmotivatie en het daadwerkelijke zoekgedrag van de werkzoekende. Meerdere schalen uit de AWV meten concepten die aansluiten bij de concepten van de Werkverkenner. Zo worden in beide instrumenten vragen gesteld over de intentie om werk te zoeken (Werkzoekintentie), datgene wat men daadwerkelijk heeft gedaan om werk te vinden (Werkzoekgedrag), het vertrouwen dat men heeft in de eigen vaardigheden om werk te vinden (Self-efficacy / Zelfvertrouwen) en hoe men de eigen kansen om werk te vinden inschat (Visie op terugkeer naar werk / Kansinschatting kans op werk). Uit validiteitsonderzoek naar de Nederlandstalige AWV, bleek daarnaast dat het mogelijk was met de AWV onderscheid te maken tussen verschillende groepen werklozen. Om deze redenen leek de AWV een geschikte vragenlijst voor het soortgenotenonderzoek. In par. 7.4.1.1 zal nader ingegaan worden op de verwachte samenhang tussen de concepten/schalen van de Werkverkenner en de AWV.

Voor het bepalen van de convergente validiteit zal een aantal werkzoekenden naast de Werkverkenner ook de AWV invullen. Aan de hand van de scores op beide vragenlijsten wordt de samenhang tussen de schalen vastgesteld. Het soortgenotenonderzoek richt zich dus niet op de individuele items die in de Werkverkenner voorkomen. Op basis van een inhoudelijke vergelijking zijn van tevoren verwachtingen geformuleerd over de verbanden tussen beide vragenlijsten. In de volgende paragraaf zullen we nader ingaan op de AWV.

---

<sup>34</sup>De afkorting COTAN betekent Commissie van Testaangelegenheden Nederland.

### 7.2.3.1 Methode

#### *Arbeidsmotivatie en Werkzoekgedrag Vragenlijst*

De AWW meet arbeidsmotivatie en het daadwerkelijke zoekgedrag van de werkzoekende. Uit onderzoek van NOA binnen gemeentelijke re-integratietrajecten uit 2008 bleek dat er op alle schalen van de AWW significante verschillen werden gevonden tussen klantprofielen, die zich onderscheidten in de afstand tot de arbeidsmarkt en de benodigde ondersteuning bij re-integratie. De AWW bestaat uit vijf schalen (zie tabel 7.13) waarvan de relatie met het vinden van werk is aangetoond. Deze schalen worden gemeten met in totaal 37 items (zie bijlage J), die worden gescoord op een vijfpuntsschaal. In tabel 7.13 staat de betrouwbaarheid per schaal zoals weergegeven in de handleiding van de AWW; deze varieert van betrouwbaar tot zeer betrouwbaar. De AWW is geschikt voor mannen en vrouwen vanaf ongeveer 16 jaar.

Tabel 7.13: Schalen van de AWW en hun betrouwbaarheid volgens de handleiding.

Schalen van de AWW	Betrouwbaarheid (Cronbach's alfa)
Belang van werk	0,84
Zelfvertrouwen	0,75
Werkzoekintentie	0,88
Werkzoekgedrag	0,90
Kans inschatting kans op werk	0,85

#### *Steekproef, onderzoeksopzet en respons*

Het streven van het soortgenotenonderzoek was om bij minimaal 300 werkzoekenden zowel de Werkverkenner als de AWW af te nemen. Alle werkzoekenden van 11 koplopervestigingen (regionaal over Nederland verspreid) werden benaderd om deel te nemen aan het soortgenoten onderzoek. Tijdens het opstarten van het onderzoek bleek het niet mogelijk om beide vragenlijsten op hetzelfde moment aan te bieden en daarbij de volgorde af te wisselen tussen de Werkverkenner en de AWW. Alle deelnemers kregen eerst (op vrijdag) de Werkverkenner aangeboden in hun werkmap en vier dagen later (op dinsdag) de AWW. De uitnodigingen werden op twee momenten verstuurd: een deel van de werkzoekenden kreeg op 9 januari 2015 een uitnodiging en een ander deel op 16 januari 2015. In totaal is aan 3014 werkzoekenden een uitnodiging verstuurd om ook de AWW in te vullen. De deelnemers kregen zowel voor de Werkverkenner als voor de AWW twee (voor de groep van 16 januari) tot drie (voor de groep van 9 januari) weken de tijd om ze in te vullen. Als de Werkzoekende de Werkverkenner nog niet had ingevuld voordat de AWW werd aangeboden, kon de werkzoekende zelf bepalen welke vragenlijst als eerste werd ingevuld. Er zijn 1015 respondenten gestart met deelname aan het soortgenotenonderzoek. Er zijn 390 werkzoekenden die gedurende de looptijd van het onderzoek ervoor kozen wel de AWW in te vullen, maar niet de Werkverkenner. Zes werkzoekenden vulden de Werkverkenner in, maar niet de AWW. In totaal 619 werkzoekenden vulden beide vragenlijsten in.

#### *Statistische analyses*

Om na te gaan of de werkzoekenden in het soortgenotenbestand representatief zijn voor de gehele populatie werkzoekenden, toetsen we of de demografische gegevens in de steekproef overeenkomen met de achtergrondgegevens van de populatie. Dit doen we op dezelfde manier als beschreven in par. 7.2.2.

Vervolgens berekenen we de gemiddelde scores op de schalen van de AWW. Daarnaast bepalen we de betrouwbaarheid van de schalen van de AWW. Dit doen we aan de hand van

Cronbach's alfa ( $\alpha$ ). We vergelijken de geobserveerde waarden van Cronbach's  $\alpha$  met de gerapporteerde waarden in de handleiding van de AWV.

We berekenen tevens de gemiddelde scores op de schalen van de Werkverkenner in de huidige onderzoeksgroep en toetsen of deze scores overeenkomen met de scores in het proefbestand (i.e. de populatie). In bijlage A staat een overzicht van de schalen van de Werkverkenner met bijbehorende items. Ook hier gebruiken we een *one sample t test* om te testen of de gevonden gemiddelde score op een schaal verschilt van de gemiddelde score in de populatie en berekenen we effectgroottes. Daarnaast bekijken we of er mogelijk sprake is van plafond- en/of bodemeffecten. We spreken van een plafondeffect als vrijwel alle werkzoekenden een hoge score hebben op een schaal. Omgekeerd spreken we van een bodemeffect als bijna alle werkzoekenden een lage score hebben op een schaal. Indien er sprake is van plafond- of bodemeffecten, dan zal de samenhang tussen schalen 'gedrukt' zijn, dat wil zeggen, lager uitvallen.

#### *Verwachte samenhang tussen de vragenlijsten*

Op basis van een inhoudelijke vergelijking van de beide vragenlijsten, waarbij gekeken is naar de onderliggende principes, definities van schalen en de operationalisaties in de vorm van specifieke vragen/schalen, zijn verwachtingen over de samenhang tussen de Werkverkenner en de AWV geformuleerd, zie tabel 7.14. Hieronder worden de verwachtingen ten aanzien van de samenhang tussen de schalen uit de Werkverkenner en AWV toegelicht.

De schaal Visie op terugkeer naar werk uit de Werkverkenner meet hoe positief/negatief men aankijkt tegen zijn/haar perspectief op de arbeidsmarkt en of men verwacht dat het lang zal duren om een baan te vinden. In de schaal Inschatting kans op werk van de AWV wordt gevraagd hoe groot men de kans inschat binnen afzienbare termijn een (passende) baan te vinden. Beide schalen vragen dus naar een inschatting van de arbeidsmarkt in relatie tot de eigen kansen om een baan te vinden. Tussen deze schalen wordt daarom een duidelijke samenhang verwacht.

De schaal Werkzoekintentie van de Werkverkenner vraagt werkzoekenden aan te geven in hoeverre zij de intentie hebben om de komende tijd werk te gaan zoeken. Met het onderdeel Werkzoekintentie uit AWV wordt gemeten hoeveel tijd de werkzoekende van plan is om de komende drie maanden te besteden aan activiteiten om werk te zoeken. Hierbij gaat het om verschillende activiteiten om werk te zoeken. Beide schalen vinden hun oorsprong in het model van de Theory of Planned Behavior (TPB) en daar uit afgeleide schalen (Van Hooft, 2004; Brouwer et al., 2011). Beide schalen hebben een gemeenschappelijke conceptuele basis en vertonen parallellen in operationalisatie, maar verschillen in de exacte vraagstelling. Op basis daarvan wordt een redelijke samenhang tussen de scores op beide schalen verwacht. Ook de schaal Zelfvertrouwen van de AWV bevat een aantal vragen die naar intentie vragen (*Ik weet precies hoe ik naar werk ga zoeken, Ik weet precies wanneer ik ga beginnen met solliciteren, Ik heb nog geen plannen om werk te zoeken, Ik heb alleen vage ideeën over hoe ik werk ga zoeken*). Daarom wordt ook met deze vragen, die in dit soortgenotenonderzoek worden samengevoegd tot de subschaal Implementatie intenties (zie hieronder voor een toelichting) een duidelijke samenhang met Werkzoekintentie van de Werkverkenner verwacht. Deze samenvoeging is gedaan op advies van de begeleidingscommissie.

De AWV-schaal Zelfvertrouwen meet meer dan alleen Zelfvertrouwen en is daarom onderverdeeld in drie subschalen, namelijk Zelfvertrouwen (*Ik denk dat ik genoeg vaardigheden heb om werk te vinden, Ik kan werkgevers overtuigen dat ik geschikt ben voor*

een baan), Arbeidskwalificaties opleiding en werkervaring (*Mijn werkervaring om een geschikte baan te vinden is, Mijn (voor)opleiding om een geschikte baan te vinden is*) en Implementatie-intenties (*Ik weet precies hoe ik naar werk ga zoeken, Ik weet precies wanneer ik ga beginnen met solliciteren, Ik heb nog geen plannen om werk te zoeken, Ik heb alleen vage ideeën over hoe ik werk ga zoeken*). Deze vier laatstgenoemde items meten in hoeverre mensen hebben nagedacht over de concretisering (“implementatie”) van hun intenties, dus in welke mate zij hun intenties omgezet hebben in concrete plannen. In de wetenschappelijke literatuur wordt dit “implementatie-intenties” genoemd (Van Hooft et al., 2005).

De schaal Werkzoekgedrag m.b.t. direct contact werkgevers uit de Werkverkenner vraagt naar wat men de afgelopen tijd heeft gedaan om werk te zoeken. Het gaat hierbij specifiek om het contact met werkgevers. De schaal Werkzoekgedrag uit de AWV meet of de kandidaat de afgelopen drie maanden tijd heeft besteed aan het zoeken naar werk. Het gaat hierbij om verschillende activiteiten om werk te zoeken. Net als bij Werkzoekintentie vinden beide schalen hun oorsprong in het TPB model en daar van afgeleide schalen (Van Hooft, 2004; Brouwer et al., 2011). Alhoewel de AWV dit aspect breder meet, wordt er tussen beide schalen een duidelijke samenhang verwacht.

De schaal Self-efficacy m.b.t. sollicitatievoorbereiding uit de Werkverkenner gaat er over in hoeverre men zich in staat acht activiteiten te ondernemen die te maken hebben met het zoeken naar werk, specifiek het solliciteren. De schaal Zelfvertrouwen van de AWV meet in hoeverre de werkzoekende vertrouwen heeft in zijn/haar vaardigheden om werk te zoeken en te vinden. De subschaal Zelfvertrouwen heeft met het solliciteren te maken. (*Ik denk dat ik genoeg vaardigheden heb om werk te vinden, Ik kan werkgevers overtuigen dat ik geschikt ben voor een baan*). Voor het soortgenoten onderzoek worden deze vragen samengenomen als subschaal Zelfvertrouwen. In beide gevallen zijn de concepten afgeleid van het begrip self-efficacy in het TPB model (Van Hooft, 2004; Eden en Aviram 1993; Sherer et al. 1992). De verwachting is dat Self-efficacy uit de Werkverkenner en de subschaal Zelfvertrouwen uit de AWV een duidelijke samenhang zullen laten zien.

Externe variabele attributie uit de Werkverkenner gaat over de mate waarin men de oorzaken voor het slagen of mislukken in het vinden van een baan verklaart door externe oorzaken. Dit concept is gebaseerd op het motivatiemodel (Brouwer et al., 2011; Gurney, 1981; Furnham en Rawles, 1996). De (sub)schalen Zelfvertrouwen en Arbeidskwalificaties opleiding en werkervaring van de AWV vragen naar het vertrouwen in de eigen vaardigheden om een baan te vinden. Het is de verwachting dat werkzoekenden die veel vertrouwen in hun eigen vaardigheden hebben, minder geneigd zijn de kans op een baan toe te schrijven aan factoren buiten zichzelf. Om die reden wordt er een omgekeerd verband tussen de scores op deze schalen verwacht. Dit verband is naar verwachting minder eenduidig dan voor de schalen hierboven, waardoor een minder hoge (omgekeerde) samenhang wordt verwacht.

Ook tussen de schalen Werkzoekattitude m.b.t. nut en noodzaak (Werkverkenner) en Belang van werk (AWV) wordt een verband verwacht. De schaal werkzoekattitude m.b.t. nut en noodzaak meet de mate waarin men een baan wil hebben, omdat men dit nuttig en/of noodzakelijk vindt. Conceptueel gezien is deze schaal afgeleid van het motivatiemodel (Bohner en Wänke, 2002). De schaal Belang van werk van de AWV meet in hoeverre de werkzoekende belang hecht aan het vinden van (betaald) werk. Naar aanleiding van het advies van de begeleidingscommissie is ook de laatste schaal aangepast voor het



soortgenoten onderzoek. De reden hiervoor is dat de schaal Belang van werk behalve aspecten van Werkzoekattitude (Van Hooft, 2004), ook het concept Subjectieve norm, oftewel de mening van anderen omvat. Dit laatste concept hebben we voor het huidige soortgenoten onderzoek buiten beschouwing gelaten, waardoor de AWW-schaal Belang van werk nu uit twee items bestaat (*Hoe belangrijk is het voor u om een dagelijks ritme in uw leven te krijgen, Hoe belangrijk vindt u het om werk te gaan zoeken*). Het is de verwachting dat er tussen de schaal Werkzoekattitude m.b.t. nut en noodzaak en de subschaal Belang van werk (AWV) een (redelijk) verband wordt gevonden. Dit verband is waarschijnlijk minder eenduidig dan tussen de Werkzoekintentie-schalen uit beide vragenlijsten, maar voor beide verbanden verwachten we een redelijke samenhang.

De schaal werkzoekattitude m.b.t. voordeel en aangenaamheid uit de Werkverkenner meet de mate waarin men een baan wil hebben, omdat men dit voordelig en/of aangenaam vindt. De schaal is afgeleid van het motivatiemodel (Bohner en Wänke, 2002). In de subschaal Belang van werk uit de AWW komt dit aspect aan bod, maar minder eenduidig. Het verwachte verband met de schaal Attitude m.b.t. nut en noodzaak is daarom gering.

Tabel 7.14: Verwachte samenhang tussen schalen Werkverkenner en AWW, op volgorde van zekerheid.

Schaal Werkverkenner	Schaal AWW	Richting van de samenhang
<i>Hoge samenhang:</i>		
Visie op terugkeer naar werk	Inschatting kans op werk	Positief
Werkzoekintentie	Implementatie-intenties (subschaal)	Positief
Werkzoekgedrag: direct contact werkgevers en sollicitaties versturen	Werkzoekgedrag	Positief
Self-efficacy m.b.t. sollicitatievoorbereiding	Zelfvertrouwen (subschaal)	Positief
<i>Redelijke samenhang:</i>		
Werkzoekintentie	Werkzoekintentie	Positief
Externe variabele attributie	Zelfvertrouwen (subschaal)	Negatief
Werkzoekattitude m.b.t. nut en noodzaak	Belang van werk (subschaal)	Positief
<i>Geen of slechts geringe samenhang:</i>		
Werkzoekattitude m.b.t. voordeel en aangenaamheid	Belang van werk	Positief
---	Arbeidskwalificaties (subschaal)	---
Acceptatiebereidheid m.b.t. onwenselijk werk	---	---
Acceptatiebereidheid m.b.t. afwijkende tijden	---	---
Gezondheidsbeleving	---	---

Tenslotte zijn er drie schalen in de Werkverkenner waarvoor op basis van de conceptuele uitgangspunten en/of operationalisatie geen equivalent in de AWW gevonden kon worden. Dit zijn de schalen: Acceptatiebereidheid m.b.t. onwenselijk werk, Acceptatiebereidheid m.b.t. afwijkende werktijden en Gezondheidsbeleving. Er wordt nauwelijks tot geen samenhang verwacht tussen deze schalen van de Werkverkenner en de schalen van de AWW. Dit wordt ook wel *divergente validiteit* genoemd, waarbij de veronderstelling is dat schalen nauwelijks samenhangen met schalen die niet hetzelfde begrip meten. Ook divergente validiteit bepalen we aan de hand van de correlaties.



We verwachten over het algemeen een positieve samenhang tussen de schalen: Een hoge score op de schaal van de Werkverkenner hangt samen met een hoge score op de schaal van de AWV. Eén uitzondering geldt hierbij voor de samenhang tussen Externe variabele attributie (Werkverkenner) en Zelfvertrouwen (AWV); hier verwachten we dat een hoge score op Externe variabele attributie samenhangt met een lage score op Zelfvertrouwen (negatieve samenhang). Om de samenhang tussen de schalen van de twee vragenlijsten te bepalen, berekenen we correlaties (Pearson's  $r$ ) tussen de schalen. Waarden van Pearson's  $r$  kunnen tussen de 0,00 en |1,00| liggen, waarbij een hoge waarde wijst op een hoge samenhang. Er zijn geen expliciete criteria voor de samenhang als het gaat om convergente validiteit. In overleg met NOA hanteren we de volgende vuistregel: Een waarde tussen 0,10 en 0,30 duidt op een lage samenhang, tussen 0,30 en 0,60 op een redelijke samenhang en waarden  $\geq 0,60$  op een hoge samenhang.

### 7.2.3.2 Resultaten

#### Representativiteit van de onderzoeksgroep

**Conclusie:** De werkzoekenden die hebben meegedaan aan het soortgenotenonderzoek zijn over het geheel genomen representatief voor de populatie.

Er zijn 619 werkzoekenden die zowel de Werkverkenner als de AWV hebben ingevuld. De minimale respons van 300 werkzoekenden is dus ruimschoots gehaald. De achtergrondgegevens van de werkzoekenden staan weergegeven in tabel 7.15.

Tabel 7.15: Resultaten van de representativiteitsanalyse van het soortgenoten onderzoek. Vergelijking van de verdelingen op de achtergrondkenmerken tussen de steekproef van het soortgenotenbestand en de totale groep respondenten in het onderzoeksbestand (referentiegroep).

Variabele	Soortgenoten ( $n = 619$ )	Referentiegroep ( $N = 53238$ )	Toetsgegevens	Effectgroottes (Cramer's $V$ of Cohen's $d$ )
Werkhervatting binnen 12 mnd (%)			$\chi^2(1) = 13,88; p < 0,01^1$	$V = 0,15$
Geen werk hervat binnen 12 maanden	34,6	47,8		
Werk hervat binnen 12 maanden	52,3	52,2		
Onbekend	13,1	-		
Leeftijd (M $\pm$ SD)	44,0 $\pm$ 11,6	42,1 $\pm$ 11,6	$t(618) = 4,0; p < 0,01$	$d = 0,16$
Leeftijd (%)			$\chi^2(1) = 16,06; p < 0,01$	$V = 0,09$
<i>t/m 27 jaar</i>	8,3	13,0		
28-49 jaar	51,7	56,2		
50-54 jaar	14,4	12,7		
$\geq 55$ jaar	25,6	18,0		
Geslacht (% vrouw)	38,0	50,9	$\chi^2(1) = 41,4; p < 0,01$	$V = 0,26$
Opleiding (%)			$\chi^2(3) = 47,0; p < 0,01^1$	$V = 0,20$
Laag	26,7	20,0		
Midden	55,7	51,3		
Hoog	16,6	28,5		
Onbekend	1,0	-		
WIA 35-min (%)	3,9	2,0	$\chi^2(1) = 11,1; p < 0,01$	$V = 0,13$
Nationaliteit			$\chi^2(3) = 10,9; p = 0,01^1$	$V = 0,08$

Variabele	Soortgenoten (n = 619)	Referentiegroep (N = 53238)	Toetsgegevens	Effectgroottes (Cramer's V of Cohen's d)
<i>Nederlandse 1e nationaliteit en geen 2e nationaliteit of een westerse 2e nationaliteit (muv Pools)</i>	82,7	95,6		
<i>Poolse 1e of 2e nationaliteit</i>	2,4	1,3		
<i>Niet-westerse 1e of 2e nationaliteit (grotendeels Turks of Marokkaans)</i>	0,5	1,1		
<i>Westerse 1e nationaliteit anders dan Nederlandse</i>	1,3	1,8		
<i>Onbekend</i>	13,1	0,1		

M = gemiddelde, SD = standaard deviatie, dikgedrukt duidt op een significant toetsresultaat ( $p < 0,05$ ).

<sup>1</sup>. De categorie 'onbekend' is niet meegenomen in de toets

In bovenstaande tabel zijn ook de referentiewaarden opgenomen. Voor alle achtergrondkenmerken vinden we significante verschillen tussen de groep werkzoekenden en de referentiewaarden. De meeste effectgroottes zijn echter klein. Dit betekent dat over het algemeen genomen, de verdelingen van de percentages per factor weinig verschillen tussen het soortgenoten-bestand en het totale onderzoekbestand. Bekijken we tabel 7.15 wat meer in detail dan zien we dat de effectgroottes voor Leeftijd en Nationaliteit zeer klein zijn (i.e. Cramer's  $V < 0,1$ ; Cohen's  $d < 0,20$ ). De effectgroottes voor de andere achtergrondkenmerken zijn klein (een waarde van Cramer's  $V$  tussen 0,1 en 0,3), met uitzondering van de factor Geslacht; Deze factor laat (afgerond) een middelgroot effect zien. Het percentage vrouwen (38,0%) is lager in het soortgenotenbestand lager dan in de referentiegroep (50,9%). Aangezien er in de vergelijking van de steekproef voor het soortgenoten onderzoek en de referentiegroep geen verschillen worden gevonden in de scores op de schalen (zie 7.4.2.3), is er geen reden om aan te nemen dat het lagere percentage vrouwen in het soortgenoten bestand de uitkomsten van het soortgenoten onderzoek beïnvloedt.

#### Scores op de AWV

**Conclusie:** We hebben de betrouwbaarheid van de schalen van de AWV onderzocht en vergeleken met de betrouwbaarheid zoals gerapporteerd in de handleiding van de AWV. We concluderen dat we, net als in de handleiding van de AWV, voor alle schalen een voldoende betrouwbaarheid vinden ( $\alpha > 0,70$ ). Een uitzondering geldt hierbij voor de subschaal Belang van werk, deze laat een lage betrouwbaarheid zien ( $\alpha = 0,48$ ). Resultaten van deze subschaal dienen we met voorzichtigheid te interpreteren.

Tabel 7.16: Scores en betrouwbaarheid per schaal van de AWV.

Schaal	Gemiddelde schaalscore (SD)	Minimum- Maximum	Betrouwbaarheid Soortgenoot-bestand (Cronbach's $\alpha$ )	Betrouwbaarheid in handleiding (Cronbach's $\alpha$ )
Zelfvertrouwen	32,9 (4,3)	16 - 40	0,81	0,75
<i>Zelfvertrouwen (subschaal)</i>	7,9 (1,5)	3 – 10	0,79	-
<i>Arbeidskwalificaties (subschaal)</i>	7,9 (1,6)	2 – 10	0,73	-
<i>Implementatie- intentie (subschaal)</i>	17,0 (2,4)	8 – 20	0,75	-
Belang van werk	21,2 (2,7)	6 - 25	0,75	0,84
<i>Belang van werk (subschaal)</i>	8,7 (1,1)	2 – 10	0,48	-
<i>Subjectieve norm (subschaal)</i>	8,0 (1,4)	2 – 10	0,76	-
Kansinschatting	7,1 (1,9)	2 - 10	0,83	0,85
Werkzoekintentie	34,5 (5,4)	15 - 50	0,80	0,88
Werkzoekgedrag	32,7 (5,3)	10 - 50	0,76	0,90

SD= standaard deviatie

De scores van de 619 respondenten op de schalen van de AWV zijn weergegeven in tabel 7.16. Voor de schaal Zelfvertrouwen is de gevonden betrouwbaarheid iets hoger dan in de handleiding van de AWV wordt vermeld ( $\alpha = 0,81$  versus  $\alpha = 0,75$ ). De betrouwbaarheid van de overige schalen zijn iets lager in de huidige onderzoeksgroep. De meeste geobserveerde betrouwbaarheden zijn van voldoende grootte ( $\alpha > 0,70$ ) om van een betrouwbare schaal te spreken met uitzondering van de subschaal Belang van werk; hier vinden we een lage betrouwbaarheid ( $\alpha = 0,48$ ). De resultaten die zijn gebaseerd op deze subschaal moeten we daarom met voorzichtigheid interpreteren.

#### Scores op de Werkverkenner

**Conclusie:** De werkzoekenden zijn qua scores op de schalen van de Werkverkenner representatief voor de populatie waar we uitspraken over willen doen. Een aantal schalen laat een significant verschil in scores zien tussen werkzoekenden in het soortgenotenbestand en werkzoekenden in de populatie, maar de gevonden verschillen zijn verwaarloosbaar klein.

De scores van de werkzoekenden op de schalen van de Werkverkenner zijn weergegeven in tabel 7.17. Deze scores zijn vergeleken met de referentiescores. Voor de helft van de schalen vinden we vergelijkbare scores ( $p > 0,05$ ). Een vijftal schalen laat significante verschillen zien tussen de gemiddelden in het soortgenotenbestand en de referentiewaarde. Als we echter naar de absolute verschillen en effectgroottes kijken, dan zien we dat deze (verwaarloosbaar) klein zijn.

Tabel 7.17: Vergelijking schaalscores in het soortgenotenbestand met scores in het totale onderzoeksbestand Werkverkenner (bereik van de schalen: 1-5)

Scores Werkverkenner	Soortgenootbestand		Referentiewaarde <sup>1</sup>		Toetsgegevens en effectgroottes
	Gemiddelde	SD	Gemiddelde	SD	
Visie op terugkeer naar werk	3,5	0,8	3,5	0,8	$t(618) = 0,61; d = 0,03$
Werkzoekgedrag m.b.t. direct contact werkgevers	2,0	0,6	1,9	0,6	<b><math>t(618) = 3,09; d = 0,13</math></b>
Gezondheidsbeleving	4,0	0,7	4,1	0,7	<b><math>t(618) = -2,28; d = -0,09</math></b>
Acceptatiebereidheid m.b.t. afwijkende tijden	2,8	1,3	2,7	1,2	<b><math>t(618) = 2,64; d = 0,11</math></b>
Acceptatiebereidheid m.b.t. onwenselijk werk	2,9	0,9	2,9	0,9	$t(618) = 1,47; d = 0,06$
Werkzoekintentie	4,1	0,7	4,1	0,7	<b><math>t(618) = -2,29; d = -0,10</math></b>
Werkzoekattitude m.b.t. nut en noodzaak	4,3	0,7	4,3	0,7	$t(618) = 0,79; d = 0,03$
Werkzoekattitude m.b.t. voordeel en aangenaamheid	3,3	0,8	3,3	0,8	$t(618) = 0,66; d = 0,03$
Externe variabele attributie	3,3	0,7	3,3	0,7	$t(618) = 0,62; d = 0,03$
Self-efficacy m.b.t. sollicitatievoorbereiding	3,8	0,6	3,9	0,6	<b><math>t(618) = -2,97; d = -0,12</math></b>

1. De referentiewaarden zijn de gegevens uit het totale onderzoeksbestand ( $N = 53238$ ); SD = standaard deviatie; dikgedrukt duidt op een significant toetsresultaat ( $p < 0,05$ ).

### Samenhang tussen de schalen (convergente validiteit en divergente validiteit)

**Conclusie:** Zoals verondersteld vinden we de richting (positief/negatief) van de verwachte samenhang terug tussen de schalen. Om te spreken van een goede convergente validiteit moeten de verwachte relaties ook de meest prominente relaties zijn. Dat wil zeggen dat een schaal van de Werkverkenner ook daadwerkelijk de hoogste correlatie heeft met de AWV-schaal die we verondersteld hadden. Dit zien we terug voor de Werkverkenner-schalen Visie op terugkeer naar werk, Werkzoekintentie en Werkzoekattitude m.b.t. nut en noodzaak. Voor de overige relaties is minder duidelijk dat de verwachte relaties ook de meest prominente zijn. Voor de meeste schalen van de Werkverkenner liggen de correlaties met schalen van de AWV lager dan we van te voren hadden bedacht. We vinden aanwijzingen voor divergente validiteit voor de schalen Werkzoekattitude m.b.t. voordeel- en aangenaamheid, Acceptatiebereidheid m.b.t. onwenselijk werk en Acceptatiebereidheid m.b.t. afwijkende werktijden; deze items hangen niet tot nauwelijks samen met schalen uit de AWV.

### Richting van de samenhang

In tabel 7.18 is de samenhang tussen alle schalen van de Werkverkenner en de AWV weergegeven. De verwachte samenhang zoals besproken in par. 7.4.1.4 is weergegeven in blauw. Zoals verondersteld vinden we voor de verwachte samenhang positieve correlaties en een verwachte negatieve correlatie tussen Externe variabele attributie en Zelfvertrouwen. Alle verwachte correlaties verschillen significant van 0 (dikgedrukt), wat betekent dat er samenhang is tussen de schalen.

### Hoogte van de samenhang

Om te spreken van een goede convergente validiteit moeten de verwachte relaties ook de meest prominente relaties zijn. In tabel 7.18 is daarom voor iedere schaal van de Werkverkenner in rood de hoogste samenhang met een AWV-schaal aangegeven. Tussen

de schalen Visie op terugkeer naar werk (Werkverkenner) en Kansinschatting kans op werk (AWV) en tussen Werkzoekintentie (Werkverkenner) en Implementatie-intenties (AWV) zien we dat de verwachte samenhang (blauwe cellen) ook de meest prominente samenhang is (rode getallen). Dit geldt ook voor de samenhang tussen Werkzoekattitude m.b.t. nut en noodzaak (Werkverkenner) en Belang van werk (AWV;  $r = 0,32$ ), hoewel de samenhang met Implementatie-intenties (AWV) in dezelfde orde van grootte is en net iets hoger uitvalt ( $r = 0,34$ ).

Voor de overige relaties is minder duidelijk dat de verwachte relaties ook de meest prominente zijn. Zo blijkt Werkzoekgedrag m.b.t. direct contact werkgevers (Werkverkenner) hoger samen te hangen met de subschaal Zelfvertrouwen (AWV;  $r = 0,36$ ) dan met de verwachte schaal Werkzoekgedrag (AWV,  $r = 0,30$ ), al is het verschil in correlaties klein. Een reden hiervoor kan zijn dat Werkzoekgedrag in de AWV vraagt naar de tijd dat men heeft besteed aan werkzoekgedrag, terwijl Werkzoekgedrag m.b.t. direct contact werkgevers in de Werkverkenner gaat over specifieke activiteiten gerelateerd aan contact met werkgevers die men de afgelopen tijd heeft uitgevoerd. Daarnaast beslaat de schaal Werkzoekgedrag m.b.t. direct contact werkgevers in de Werkverkenner een specifiek deel van werkzoekgedrag. De Werkverkenner bevat ook een item Werkzoekgedrag m.b.t. sollicitaties versturen waarbij wordt gevraagd of iemand (digitaal) sollicitaties heeft verzonden (zie bijlage A). Dit item is echter geen onderdeel van de schaal Werkzoekgedrag m.b.t. direct contact werkgevers. Beide aspecten van werkzoekgedrag komen uit de Job Search Behavior Index (Blau, 1994).

Self-efficacy m.b.t. sollicitatievoorbereiding (Werkverkenner) heeft de hoogste correlatie met de subschaal Implementatie-intenties (AWV;  $r = 0,44$ ), terwijl we – op advies van de begeleidingscommissie – verwachtten dat de hoogste samenhang met de subschaal Zelfvertrouwen (AWV;  $r = 0,30$ ) zou zijn. Voor Externe variabele attributie (Werkverkenner) vinden we de hoogste samenhang ( $r = -0,36$ ) met Inschatting kans op werk (AWV), terwijl we de hoogste samenhang verwachtten met de subschaal Zelfvertrouwen (AWV;  $r = -0,18$ ). We waren over deze verwachte samenhang echter al minder zeker, zoals ook in tabel 7.14 was aangegeven (i.e. redelijke samenhang). Externe variabele attributie meet het toeschrijven van het slagen of falen van het eigen gedrag aan oorzaken die buiten de persoon zelf liggen. Iemand die hoog scoort op deze factor verwacht in redelijke mate dat de kans op werk laag zal zijn (negatieve samenhang).

Alleen tussen de schalen Visie op terugkeer naar werk (Werkverkenner) en Kansinschatting kans op werk (AWV) vinden we een hoge samenhang daar waar we ook een hoge samenhang verwachtten ( $r \geq 0,60$ ). Voor de overige schalen in de Werkverkenner liggen de correlaties met schalen van de AWV lager dan we van tevoren hadden bedacht. Zo vinden we voor Werkzoekintentie (Werkverkenner) een redelijke samenhang ( $r = 0,38$ ) met de AWV-subschaal Implementatie-intenties, terwijl we een hoge samenhang hadden verwacht. Voor de schaal Werkzoekintentie (Werkverkenner) vinden we een lage samenhang met Werkzoekintentie uit de AWV, daar waar we een redelijke samenhang hadden verwacht. Een mogelijke verklaring voor de lagere samenhang betreft de verschillen in vraagstellingen en antwoordmogelijkheden tussen de AWV en de Werkverkenner. Werkzoekintentie in de AWV, bijvoorbeeld, wordt aan de hand van een aantal items gemeten waarbij de vraag wordt gesteld 'hoeveel tijd wilt/denkt u de komende periode te besteden aan...'. Deze items kunnen beantwoord worden met een categorie tussen geen tijd (0 punten) tot zeer veel tijd (5 punten). Items in de Werkverkenner die Werkzoekintentie meten zijn bijvoorbeeld 'Ik ben van plan om er de komende maand alles aan te doen om werk te vinden' en 'Ik weet hoe ik de komende maand kan proberen werk te vinden' en worden op een schaal beantwoord tussen zeer mee oneens (0 punten) en zeer mee eens (4 punten). Mogelijk zorgt de

combinatie van vraagstelling en bijbehorende antwoordschaal ervoor dat ogenschijnlijk verwante items toch minder samenhang vertonen dan verwacht, wat een nadelige invloed heeft op de convergente validiteit. Daarnaast kan het zijn dat de tijd tussen het invullen van de Werkverkenner en de AWV van invloed is op de samenhang. De Werkverkenner en de AWV werden niet op dezelfde dag aangeboden en werkzoekenden konden zelf bepalen wanneer ze welke vragenlijst invulden. Meer tijd tussen de twee invulmomenten kan een negatieve invloed hebben op de samenhang. Wanneer bijvoorbeeld een werkzoekende een van de vragenlijsten heeft ingevuld en vervolgens in de daarop volgende dagen of weken een reeks afwijzingen op sollicitaties te verduren krijgt, dan valt te verwachten dat deze afwijzingen invloed kunnen hebben op het invullen van de andere vragenlijst. Wanneer de vragenlijsten op dezelfde dag worden ingevuld, dan blijft de invloed van dergelijke factoren beperkt. Hoe langer de tijd tussen het invullen van de vragenlijsten, hoe groter de kans op gebeurtenissen die een invloed kunnen hebben.

Daarnaast wordt er veelal bij (diagnostische) testen vanuit gegaan dat begrippen die worden gemeten, zoals intenties en attitudes, stabiel zijn over tijd. Intenties en attitudes zijn echter helemaal niet stabiel en kunnen veranderen. Aangezien de Werkverkenner en de AWV niet op dezelfde dag zijn ingevuld, kan dit (negatieve) gevolgen hebben voor de samenhang tussen de schalen uit de Werkverkenner en de AWV.

Naast convergente validiteit kunnen we ook kijken naar 'divergente validiteit'. We spreken van een goede divergente validiteit als schalen niet samenhangen met schalen die iets heel anders beogen te meten. Voor de schaal Werkzoekattitude m.b.t. voordeel en aangenaamheid uit de Werkverkenner was de hypothese dat er een (zeer) geringe samenhang zou zijn met de schaal Belang van werk uit de AWV. Ook voor de Werkverkenner-schalen Acceptatiebereidheid m.b.t. Onwenselijk werk, Acceptatiebereidheid m.b.t. afwijkende werktijden verwachtten we geen relatie met een schaal van de AWV (divergente validiteit). We vinden aanwijzingen voor divergente validiteit voor de items Werkzoekattitude m.b.t. nut en noodzaak, Acceptatiebereidheid m.b.t. onwenselijk werk en Acceptatiebereidheid m.b.t. afwijkende tijden; deze schalen hangen nauwelijks samen met schalen uit de AWV. We hebben bewijs gevonden voor divergente validiteit van de Werkverkenner, aangezien voor alle drie de schalen een lage tot geen samenhang met schalen van de AWV is gevonden. Ook voor de schaal Gezondheidsbeleving uit de Werkverkenner hadden we geen samenhang met een schaal uit de AWV verondersteld, maar voor deze schaal vinden we echter een lage tot redelijke samenhang met schalen van de AWV. Zo zien we een redelijke samenhang met Inschatting kans op werk ( $r = 0,36$ ), de subschaal Zelfvertrouwen ( $r = 0,30$ ) en de subschaal Implementatie-intenties ( $r = 0,30$ ) uit de AWV.

#### *Plafond- of bodemeffecten*

**Conclusie:** We sluiten uit dat plafond- en bodemeffecten een rol spelen in de gevonden correlaties tussen de schalen van de Werkverkenner en de AWV.

Omdat de gevonden correlaties niet hoog zijn, hebben we gekeken of er plafond- en/of bodemeffecten te vinden zijn in de items van de Werkverkenner en de AWV. In de rapportage van het proefbestand van de Werkverkenner hebben we al geconcludeerd dat er geen sprake is van plafond- en bodemeffecten. Ook voor de scores op de AWV vinden we geen indicatie voor plafond- of bodemeffecten. We hebben daarom geen redenen om aan te nemen dat plafond- of bodemeffecten een rol spelen in de hoogte van de gevonden correlaties.

Tabel 7.18: Correlaties (Pearson's  $r$ ) tussen alle schalen van de Werkverkenner en de AWV (met in blauw de schalen waarbij samenhang verondersteld wordt)\*.

AWV \ Werkverkenner	Belang van werk <sup>1</sup>	Inschatting kans op werk	Werkzoekgedrag	Werkzoekintentie	Zelfvertrouwen <sup>2</sup>	Arbeidskwalificaties werk en opleiding <sup>3</sup>	Implementatie-intentie <sup>4</sup>
Visie op terugkeer naar werk	0,17	0,69	0,06	0,01	0,39	0,28	0,29
Werkzoekintentie	0,25	0,21	0,21	0,23	0,29	0,20	0,38
Werkzoekgedrag m.b.t. contact werkgevers	0,21	0,34	0,30	0,17	0,36	0,30	0,24
Self-efficacy m.b.t. sollicitatievoorbereiding	0,16	0,16	0,17	0,17	0,30	0,26	0,44
Externe variabele attributie	-0,05	-0,36	0,05	0,05	-0,18	-0,08	-0,16
Werkzoekattitude m.b.t. nut en noodzaak	0,32	0,14	0,16	0,19	0,19	0,15	0,34
Werkzoekattitude m.b.t. voordeel- en aangenaamheid	0,16	0,12	0,16	0,14	0,17	0,09	0,16
Acceptatiebereidheid m.b.t. onwenselijk werk	0,08	-0,05	0,08	0,11	-0,05	-0,04	0,10
Acceptatiebereidheid m.b.t. afwijkende werktijden	0,15	0,12	0,05	0,06	0,09	0,05	0,06
Gezondheidsbeleving	0,22	0,36	0,13	0,12	0,33	0,32	0,33

\* Dikgedrukt duidt op een significant toetsresultaat ( $p < .05$ ), blauw gearceerde cellen geven de verwachte samenhang weer (donkerblauw: hoge verwachte samenhang, lichtblauw: redelijke verwachte samenhang), rode getallen duiden de hoogste samenhang met een AWV-schaal aan voor de betreffende Werkverkenner-schaal.

1. Het gaat hier om de subschaal Belang van werk (dus niet de originele schaal).
2. Het gaat hier om de subschaal Zelfvertrouwen (dus niet de originele schaal).
3. De items die horen bij Arbeidskwalificaties komen uit de oorspronkelijke Zelfvertrouwen-schaal.
4. De items die horen bij Implementatie-intenties komen uit de oorspronkelijke Zelfvertrouwen-schaal.

### 7.3 Sociaal wenselijke beantwoording

**Conclusie:** Uit de verdeling van de antwoorden op de items vinden we geen indicatie voor een hoge mate van sociaalwenselijke beantwoording. We concluderen dat er geen verdere actie nodig is (zoals bijvoorbeeld herformulering van de vragen).

De Werkverkenner heeft, net als elk zelfbeoordelingsinstrument, het risico om sociaalwenselijke antwoorden te genereren. Er bestaan methoden die speciaal ontwikkeld zijn om de kans te achterhalen dat bepaalde personen sociaalwenselijke antwoorden geven. Hiervoor moeten werkzoekenden additionele vragen invullen (die niets met het zoeken naar werk te maken hebben) naast de Werkverkenner. Dit was echter niet mogelijk in de huidige onderzoeksopzet. Tevens heeft de manier van afname van de vragenlijst (digitaal of schriftelijk) mogelijk invloed op de mate van sociaalwenselijke beantwoording.

De vragenlijst Werkverkenner 1.0+ werd door UWV digitaal afgenomen. Voor de ontwikkeling van Werkverkenner 1.0 is de vragenlijst destijds schriftelijk afgenomen. Een meta-analyse van Dodou en de Winter (2014) laat zien dat er gemiddeld genomen geen verschil is in sociaalwenselijke beantwoording tussen vragenlijsten die digitaal en schriftelijk werden afgenomen. De meta-analyse laat tevens zien dat de resultaten niet eenduidig zijn. De richting van de resultaten bleek afhankelijk te zijn van het instrument dat werd gebruikt om sociaalwenselijkheid te meten (bijv. de “Balanced Inventory of Desirable Responding (BIDR)” of de “Marlowe-Crowne Social Desirability Scale (MCDS)”. Met de BIDR werd gevonden dat digitale afname in vergelijking met schriftelijke afname resulteerde in meer sociaalwenselijke beantwoording, terwijl met de MCDS digitale afname resulteerde in minder sociaalwenselijke beantwoording.

Verder is een interessante bevinding in de literatuur dat het rapporteren van risicogedrag (bijv. alcoholgebruik of onveilig rijden) frequenter wordt gerapporteerd bij digitale afname dan bij schriftelijke afname (Booth-Kewley, Larson, & Miyoshi, 2007).

Het is niet uit te sluiten dat op sommige vragen sociaalwenselijke antwoorden gegeven worden. Als er sterke vermoedens bestaan voor sociaalwenselijke beantwoording dient de vraag anders geformuleerd te worden. Een geringe mate van sociaalwenselijke beantwoording is op zich niet erg voor het bepalen van een goede voorspelregel voor status werkhervatting (hoofdstukken 3 en 4). Het is aannemelijk dat de mate van sociaalwenselijkheid in de antwoorden van de huidige steekproef niet zal verschillen van de mate van sociaalwenselijkheid van toekomstige werkzoekenden. Daarom zal de kwaliteit van de voorspelregel er niet door verminderen. Om toch, gegeven de beperkingen van de huidige onderzoeksopzet, een indruk te krijgen of er veel sociaalwenselijke antwoorden voorkomen, hebben we beschrijvende analyses uitgevoerd. We kijken naar de verdeling van de antwoorden per item, om te zien of bijvoorbeeld slechts één deel van een schaal voorkomt in de beantwoording (bijv. bij een vijfpuntsschaal: alleen een score van 3 of hoger). Deze analyses hebben we uitgevoerd met de respondenten van het proefbestand. Dit betreft werkzoekenden ( $n = 25676$ ) die zijn ingestroomd in de WW tussen december 2013 tot en met 1 juli 2014. Zij voldoen aan dezelfde criteria als genoemd in par. 2.1.

Voor de totale onderzoeksgroep van het proefbestand hebben we de verdeling bekeken van de afzonderlijke items. Er zijn geen items gevonden waarbij slechts een deel van de antwoordschaal werd gebruikt. Alle mogelijke scores werden geobserveerd.



## 7.4 Predictieve validiteit

**Conclusie:** De predictieve validiteit van Werkverkenner 2.0 is redelijk (AUC = 0,78). We kunnen met de 18 factoren van Werkverkenner 2.0 de status werkhervatting binnen 12 maanden voor 70,3% correct voorspellen.

Bij het bepalen van de predictieve validiteit van een test gaat het om de mate waarin de test een criterium voorspelt (Lord & Novick, 1968). Het criterium is zo gekozen dat het overeenkomt met het doel van de test in de beoogde populatie. Predictieve validiteit kan worden gezien als een speciale vorm van criteriumvaliditeit, waarbij het criterium is gemeten op een later tijdstip dan het tijdstip van de testafname. In ons geval is een belangrijk doel van de Werkverkenner om de kans op werkhervatting van een werkzoekende te voorspellen. We beschouwen dan ook status Werkhervatting als het belangrijkste criterium. We hebben dit criterium gemeten op drie momenten: op 6 maanden na aanvang van de WW (en dus ca. 4 maanden na invulling van de Werkverkenner), op 9 maanden na aanvang van de WW en op 12 maanden. Op basis van de voorspelde kans op werkhervatting binnen 12 maanden zal UWV de verleende services afstemmen op de betreffende werkzoekende (e.g., voor degenen met een lage kans kan UWV besluiten om aanvullende dienstverlening in te zetten om ervoor te zorgen dat zij weer het werk hervatten). De status Werkhervatting binnen 12 maanden kunnen we daarom als het belangrijkste criterium beschouwen dat voorspelt dient te worden door de Werkverkenner. We beperken ons in deze paragraaf tot het beschrijven van de predictieve validiteit van de Werkverkenner voor dit criterium.

In hoofdstuk 3 van dit rapport hebben we op basis van de factoren van de Werkverkenner (gemeten met Werkverkenner 1.0+) het meest spaarzame, interpreteerbare voorspelmodel bepaald voor status werkhervatting op 12 maanden. De 18 factoren van dit model vormen samen Werkverkenner 2.0.

De predictieve validiteit van Werkverkenner 2.0 is bepaald met gebruikmaking van een test set. De personen in deze testset zijn niet gebruikt voor het ontwikkelen van het voorspelmodel. Zodoende kan een goede schatting worden gemaakt van de voorspellende eigenschappen van Werkverkenner 2.0 voor toekomstige werkzoekenden die deze Werkverkenner gaan invullen. De belangrijkste globale voorspellende eigenschappen die we hebben onderzocht zijn de Area-Under-the-Curve (AUC) en de Brierscore. De AUC is gelijk aan 0,778, dit geeft aan dat het model een redelijk onderscheiden vermogen heeft ( $0,70 < AUC < 0,80$ ; Weinstein & Fineberg, 1980). De Brierscore is gelijk aan 0,191, dit geeft aan dat het model goed passend is (i.e. de gemiddelde voorspelfout is lager dan 0,20).

Op basis van verschillende afkapwaarden op de voorspelde kans, hebben we de volgende specifieke voorspellende eigenschappen van het model vastgesteld: het % correct voorspeld, de sensitiviteit en de specificiteit. Een afkapwaarde van 0,45 leverde het hoogste % correct voorspeld op, namelijk 70,3%. De bijbehorende sensitiviteit, d.w.z. het percentage werkzoekenden van wie terecht voorspeld wordt dat zij het werk hervatten binnen 12 maanden, is 80%. De bijbehorende specificiteit, d.w.z. het percentage werkzoekenden van wie terecht voorspeld wordt dat zij niet het werk hervatten binnen 12 maanden, is 60%. Idealiter zouden we voor de UWV praktijk een iets hogere specificiteit willen bereiken, zodat de meeste werkzoekenden die inderdaad geen werk vinden binnen 12 maanden in aanmerking komen voor aanvullende dienstverlening. We kunnen dit realiseren met Werkverkenner 2.0 door een hogere afkapwaarde op de voorspelde kans te gebruiken. Bijvoorbeeld, een afkapwaarde van 0,72 levert een specificiteit op van 90% (en een sensitiviteit van 40%). Zie hoofdstuk 3, Tabel 3.5a, voor een overzicht van de verschillende

waarden van specificiteit en sensitiviteit en de bijbehorende afkapwaarden op de voorspelde kans op werkhervatting binnen 12 maanden.

De gekozen afkapwaarde heeft consequenties voor de praktijk van UWV. Als alle werkzoekenden met een kans lager dan 0,45 aanvullende dienstverlening krijgen, dan zal UWV moeten inschatten of dit qua budget en capaciteit haalbaar is. De uiteindelijke afkapwaarde die in de praktijk van UWV gebruikt zal worden, zal daarom in overleg met verschillende partijen worden bepaald.

## 8 Voorspelmodel per klantgroep

**Conclusie:** UWV heeft binnen de totale onderzoeksgroep drie klantgroepen gedefinieerd met een grote afstand tot de arbeidsmarkt om vast te stellen of een voorspelmodel ontwikkeld voor een specifieke klantgroep beter voorspelt dan het model dat ontwikkeld is voor de totale onderzoeksgroep. Deze drie klantgroepen zijn: mensen van 50 tot en met 54 jaar, 55-plussers en mensen met een WIA 35-min indicatie. Uit de analyses blijkt dat we voor de voorspelling van werkhervatting per klantgroep kunnen volstaan met het spaarzame, interpreteerbare model dat is ontwikkeld met behulp van de totale onderzoeksgroep (van hoofdstuk 3); dit voorspelmodel hebben we gekozen als Werkverkenner 2.0 (zie hoofdstuk 4).

De voorspellende eigenschappen van Werkverkenner 2.0 voor een klantgroep zijn vrijwel gelijk aan de voorspellende eigenschappen van het model ontwikkeld voor een specifieke klantgroep. Dit betekent dat we geen betere predictie accuraatheid verkrijgen door het model afzonderlijk te ontwikkelen per klantgroep. Met andere woorden, we kunnen volstaan met de voorspelling door Werkverkenner 2.0. Aangezien het hier om kwetsbare groepen op de arbeidsmarkt gaat waarbij de meeste mensen niet binnen een jaar werk vinden is het de vraag of de naïeve methode, door te stellen dat iedereen niet binnen een jaar het werk hervat, al volstaat. Het blijkt dat in vergelijking met de naïeve methode, Werkverkenner 2.0 beter voorspelt voor de klantgroepen 50-54 jaar en 55-plus en vrijwel gelijk voorspelt voor de klantgroep met een WIA 35-min indicatie (zie onderstaande tabel).

	Totale groep	WW 50-54	WW 55+	WIA 35-min
Percentage correct voorspeld Werkverkenner 2.0	70,3	67,4	79,1	78,8
Percentage correct voorspeld klant-specifiek model	-	67,4	79,9	78,4
Percentage correct voorspeld naïeve methode	51,8	57,3	75,4	79,1

De klant-specifieke modellen geven inzicht in de factoren die belangrijk zijn voor de diagnose. De meeste factoren in de klant-specifieke modellen overlappen met Werkverkenner 2.0. De toegevoegde bijdrage van de nieuwe factoren was erg klein. We concluderen dat voor de diagnose per klantgroep de factoren van Werkverkenner 2.0 volstaan.

In dit laatste hoofdstuk werken we onderzoeksvraag 1 “Wat zijn voorspellende persoons- en situatiekenmerken voor werkhervatting van de WW-instroom?” (zie par. 1.2) op verzoek van UWV uit voor drie specifieke klantgroepen met een grote afstand tot de arbeidsmarkt:

1. WW-instroom tussen 50 en 54 jaar;
2. WW-instroom vanaf 55 jaar; en
3. WW-instroom met een WIA 35-min-indicatie<sup>35</sup>.

Het beste voorspelmodel voor de totale onderzoeksgroep is vastgesteld in hoofdstukken 3 en 4. We kozen het meest spaarzame, interpreteerbare meervoudige logistisch regressie-model als beste model. Dit model bestaat uit 18 factoren die voorspellend zijn voor

<sup>35</sup> De inclusiecriteria van de onderzoeksgroep staan vermeld in hoofdstuk 2.

werkhervatting binnen 12 maanden (par 3.5.3). De bijbehorende gewichten van de 18 factoren zijn weergegeven in tabel 3.6. We verwijzen naar dit voorspelmodel als Werkverkenner 2.0.

In dit hoofdstuk beantwoorden we de volgende deelvragen:

1. *Hoe goed kan Werkverkenner 2.0 de kans op werkhervatting voorspellen voor de klantgroepen apart (WW 50-54 jaar, WW 55 jaar en ouder, WIA 35-min-indicatie)? En levert een apart ontwikkeld (specifiek) model een betere voorspelling op voor een klantgroep vergeleken met Werkverkenner 2.0?*
2. *Wat zijn voorspellende persoons- en situatietekenen voor werkhervatting binnen 12 maanden per klantgroep?*

## 8.1 Methode per klantgroep

Om bovenstaande vragen te onderzoeken selecteerden we uit de vier bestanden die we gebruikten in hoofdstuk 3 (het totale bestand, de ontwikkelset, de validatieset en de testset) de gegevens apart voor de drie klantgroepen: 50-54 jarigen, 55 plussers en WW'ers met een WIA 35-min-indicatie (zie tabel 8.1). Per klantgroep beschikten we dus over vier datasets.

Allereerst hebben we in de testset van iedere klantgroep de voorspelde kans op werkhervatting berekend met Werkverkenner 2.0. Vervolgens bepaalden we met deze voorspelde kans en de werkelijk status werkhervatting in de testset, de waarden van de globale maten: de AUC en de Brierscore. Ook bepaalden we de specifieke voorspellende eigenschappen: de specificiteit (voor gegeven sensitiviteit) en sensitiviteit (voor gegeven specificiteit) en het percentage correcte voorspelling, op gelijke wijze als in hoofdstuk 3.

Voor het ontwikkelen van de *specifieke modellen* hebben we per klantgroep de volgende 4 stappen in de analyses doorlopen (zie ook par. 3.5 en bijlage C):

- > Stap 1. Het schatten van het meervoudige logistische regressiemodel met alle factoren van Werkverkenner 1.0+ (weergegeven in tabel 1.1) in de ontwikkelset;
- > Stap 2. Het schatten van het interpreteerbare, meervoudige model (i.e. zonder de factoren die moeilijk te interpreteren zijn in het model van stap 1) in de ontwikkelset;
- > Stap 3. Het schatten van het meest spaarzame, interpreteerbare meervoudige model (door middel van stapsgewijze selectie met de AIC en BIC methode met gebruikmaking van de ontwikkelset en de validatieset). Dit laatste model noemen we het specifieke model voor de klantgroep.
- > Stap 4. De parameters van het beste specifieke model (van Stap 3) zetten we vast en we berekenen de voorspelde kans op werkhervatting in de testset. Vervolgens bepaalden we met deze voorspelde kans en de werkelijk status werkhervatting in de testset, de waarden van de globale maten: de AUC en de Brierscore. Ook bepaalden we de specifieke voorspellende eigenschappen: de specificiteit (voor gegeven sensitiviteit) en sensitiviteit (voor gegeven specificiteit) en het percentage correcte voorspelling.

Vervolgens vergelijken we per klantgroep de voorspellende eigenschappen van het model van Werkverkenner 2.0 met die van het klant-specifieke model. Om te toetsen of de globale maat (i.e. de AUC) verschilt tussen de twee modellen, voeren we de DeLong test uit (DeLong, DeLong, & Clarke-Pearson, 1988). Als deze toets niet significant is, concluderen we dat er geen verschil is in onderscheidend vermogen tussen de twee modellen. Ondanks geen verschil in deze globale maat kunnen er kleine verschillen optreden in de specifieke voorspeleigenschappen tussen de twee modellen. Het kan bijvoorbeeld toch voorkomen dat het percentage correct iets hoger is in het ene model dan in het andere. Dit zijn toevallige verschillen en deze verschillen komen door het vooraf vastzetten van de waarden voor sensitiviteit of specificiteit.

Voor het beantwoorden van de 2<sup>e</sup> onderzoeksvraag inspecteren we de factoren die zijn geïnccludeerd in het meest spaarzame, interpreteerbare klantgroep-specifieke model. Dit model geeft inzicht in welke factoren invloed hebben op het hervatten van werk in de specifieke klantgroep (dus belangrijk kunnen zijn voor de diagnose). We vergelijken de bijdrage van de factoren aan de voorspelling (i.e.  $\Delta R^2$ ) in het klant-specifieke model met de bijdrage van de factoren in het model van Werkverkenner 2.0.

## 8.2 Resultaten per klantgroep

Voordat we met het beschrijven van de resultaten van de voorspelmodellen beginnen bekijken we eerst het percentage werkhervatters in de verschillende data sets per klantgroep en in de totale onderzoeksgroep. Per klantgroep zijn de percentages in de vier datasets vrijwel gelijk (zie tabel 8.1). Tabel 8.1 laat tevens zien dat het percentage werkhervatters binnen 12 maanden het laagst is in de WIA 35-min groep (20%).

Tabel 8.1: Aantallen en percentages van de werkzoekenden die het werk hebben hervat binnen de drie klantgroepen, verdeeld over de ontwikkel-, validatie- en testset.

	Totale bestand	Ontwikkelset	Validatieset	Testset
<b>Klantgroep WW 50 t/m 54</b>				
Geen werk hervat < 12 mnd. (n (%))	3844 (58%)	1888 (57%)	1013 (60%)	943 (57%)
Werk hervat < 12 mnd. (n (%))	2773 (42%)	1396 (43%)	674 (40%)	703 (43%)
Totaal (n (%))	6617(100%)	3284 (100%)	1687 (100%)	1646 (100%)
<b>Klantgroep WW 55-plus</b>				
Geen werk hervat < 12 mnd. (n (%))	6940 (75%)	3448 (58%)	1728 (74%)	1764 (75%)
Werk hervat < 12 mnd. (n (%))	2369 (25%)	1173 (25%)	621 (26%)	575 (25%)
Totaal (n (%))	9309(100%)	4621 (100%)	2349 (100%)	2339 (100%)
<b>Klantgroep WIA 35-min</b>				
Geen werk hervat < 12 mnd. (n (%))	858 (80%)	419 (80%)	208 (79%)	231 (79%)
Werk hervat < 12 mnd. (n (%))	218 (20%)	102 (20%)	55 (21%)	61 (21%)
Totaal (n (%))	1076 (100%)	521 (100%)	263 (100%)	292 (100%)
<b>Totale onderzoeksgroep</b>				
Geen werk hervat < 12 mnd. (n (%))	25409 (48%)	12684 (48%)	6324 (48%)	6401 (48%)
Werk hervat < 12 mnd. (n (%))	27670 (52%)	13857 (50%)	6945 (52%)	6868 (52%)
Totaal (n (%))	53079 (100%)	26541(100%)	13269(100%)	13269(100%)

We kunnen op basis van deze tabel het percentage correct voorspeld bepalen in de testset per klantgroep volgens de naïeve voorspelmethode (voor uitleg van deze methode, zie par. 3.3). Dit percentage is 57% voor de groep WW 50 t/m 54, 75% voor de groep WW 55+ en 79% voor de groep WIA 35-min. Voor deze laatste twee klantgroepen (WW 55+ en WIA 35-min) wordt het dus nog lastig de voorspelling te verbeteren met een model.

### 8.2.1 Resultaten voor de klantgroep WW 50 t/m 54 jaar

Voor de klantgroep WW 50 t/m 54 jaar zijn de voorspellende eigenschappen van Werkverkenner 2.0 (zie tabel 8.2a) vrijwel gelijk aan die van het *klantgroep-specifieke* model (zie tabel 8.2b). De globale maat (de AUC) van Werkverkenner 2.0 verschilt niet significant van de globale maat van het klant-specifieke model ( $z = -0,24$ ;  $p = 0,81$ ). De Brierscores van beide modellen zijn gelijk; ook het hoogste percentage correct voorspeld (67,4%) is gelijk in

beide modellen. Dit percentage ligt hoger dan de 57% correct voorspeld met de naïeve methode. De maximale waarde van de som van de sensitiviteit en specificiteit is met Werkverkenner 2.0 (134,1) vrijwel gelijk aan de waarde van het klantgroep-specifieke model (134,2).

Tabel 8.2a: Voorspellende eigenschappen van Werkverkenner 2.0 voor de voorspelling van **status werkhervatting binnen 12 maanden in de klantgroep WW 50 - 54**;  $n = 1646$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
20,1	95,0	0,64	63,0	115,1	210,1	0,73	0,21
31,4	90,0	0,58	65,0	121,5	211,5		
40,1	85,1	0,54	65,9	125,2	210,2		
48,7	80,0	0,50	66,6	128,6	208,6		
<b>57,3</b>	<b>75,0</b>	<b>0,46</b>	<b>67,4</b>	<b>132,3</b>	<b>207,3</b>		
63,4	70,0	0,44	67,2	133,4	203,4		
70,0	63,4	0,40	66,2	133,4	196,8		
75,0	59,2	0,37	65,9	134,1	193,3		
79,9	51,2	0,34	63,5	131,2	182,4		
85,1	44,0	0,30	61,5	129,1	173,1		
90,0	36,6	0,26	59,4	126,6	163,2		
95,0	21,0	0,20	52,6	116,0	137,0		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met de hoogste waarde voor het percentage correct voorspeld is donkerblauw.

Tabel 8.2b: Voorspellende eigenschappen van het klantgroep-specifieke model voor de voorspelling van **status werkhervatting binnen 12 maanden in de klantgroep WW 50 - 54**;  $n = 1643^1$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
21,6	95,0	0,68	63,6	116,6	211,6	0,73	0,21
30,7	90,0	0,61	64,6	120,7	210,7		
41,0	85,0	0,56	66,2	126,0	211,0		
48,8	80,0	0,52	66,6	128,8	208,8		
56,3	75,0	0,48	67,0	131,3	206,3		
<b>63,9</b>	<b>70,0</b>	<b>0,44</b>	<b>67,4</b>	<b>133,9</b>	<b>203,9</b>		
70,0	64,3	0,41	66,7	134,2	198,5		
75,0	56,7	0,37	64,5	131,7	188,4		
79,9	51,4	0,34	63,6	131,3	182,7		
85,1	44,3	0,31	61,7	129,3	173,6		
90,0	34,5	0,26	58,2	124,5	159,0		
95,0	20,5	0,20	52,4	115,6	136,1		

<sup>1</sup> De personen in de categorie vrouw getrouwd/woont samen, jongste kind thuis 0-6 jaar zijn verwijderd uit de analyses; slechts 1 persoon in de ontwikkelset en 3 personen in de testset scoorden in deze categorie. AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met de hoogste waarde voor het percentage correct voorspeld is donkerblauw.

Voor het beantwoorden van de vraag welke factoren belangrijk zijn voor deze klantgroep, inspecteren we het klantgroep-specifieke model, weergegeven in bijlage K, Tabel K.1. We hebben de factoren wederom geordend van meest belangrijk naar minst belangrijk voor de

voorspelling op grond van de relatieve toename in Nagelkerke  $R^2$  ( $\Delta R^2$ ). Het klantgroep-specifieke model voor de klantgroep WW 50 t/m 54 heeft in totaal 13 factoren. De belangrijkste 10 factoren in dit model komen ook voor in Werkverkenner 2.0 (zie tabel 8.3).

Tabel 8.3: Overzicht van de factoren in Werkverkenner 2.0 en in de klantgroep-specifieke modellen. De cijfers in de cellen geven de rangorde aan in belangrijkheid op basis van de relatieve bijdrage ( $\Delta R^2$ ).

Factor	Werkverkenner 2.0	Klantgroep-specifieke modellen		
	Totale groep	50-54	WW 55+	WIA 35-min
Leeftijd	1	10	4	3
Aantal jaren in laatste functie	2	3	1	-
Visie op terugkeer naar werk	3	5	2	-
Bemiddelingsberoep soort	4	1	3	-
Samenstelling huishouden	5	6	6	2
Maximale duur WW-recht (> 12 mnd.)	6	7	13	1
Bedrijfstak	7	4	5	-
Werkzoekgedrag m.b.t. direct contact werkgevers	8	9	8	7
Gezondheidsbeleving	9	2	9	10
Balans voor- en nadelen niet werken	10	-	10	-
WIA 35-min-indicatie	11	-	-	-
Problemen Nederlandse taal	12	-	-	8
Acceptatiebereidheid m.b.t. voltijd baan	13	8	-	-
Algemeen werkvermogen	14	-	12	-
Aantal uren in staat te werken	15	-	11	9
Opleidingsniveau	16	-	-	-
Acceptatiebereidheid m.b.t. afwijkende tijden	17	-	-	-
Aanvullende inkomsten uit werk naast WW	18	-	15	4
Aanvullende uitkering	-	11	-	-
Fractie zelf ontslag genomen	-	12	-	-
Werkzoekintentie	-	13	-	-
Regio koplopervestiging	-	-	7	-
Acceptatiebereidheid m.b.t. onwenselijk werk	-	-	-	-
Financiële noodzaak om < 2 mnd. werk te vinden	-	-	14	5
Subjectieve norm familie en partner	-	-	-	-
Bemiddelingsberoep: niveau	-	-	-	6
Belemmering: financiële problemen	-	-	-	11
Percentage correct voorspeld Werkverkenner 2.0	70,3	67,4	79,1	78,8
Percentage correct voorspeld klant-specifiek model	-	67,4	79,9	78,4
Percentage correct voorspeld naïeve methode	52,1	57,3	75,4	79,1

Omdat we een specifieke leeftijdsgroep hebben geselecteerd (50 t/m 54 jarigen) is de factor leeftijd minder belangrijk in dit model en komt pas op de 10<sup>e</sup> plaats (tabel 8.3). Nieuwe factoren voor deze klantgroep (t.o.v. de totale onderzoeksgroep) zijn: Aanvullende uitkeringen, Fractie zelf ontslag genomen en Werkzoekintentie. De relatieve bijdrage van ieder van deze nieuwe factoren is echter erg laag (i.e.  $\Delta R^2 \leq 0,002$ ; bijlage K tabel K.1). We



concluderen hieruit dat de toegevoegde waarde voor de diagnose van deze nieuwe factoren (t.o.v. de factoren van Werkverkenner 2.0) klein is.

## 8.2.2 Resultaten voor de klantgroep WW 55+

Tabel 8.4a: Voorspellende eigenschappen van Werkverkenner 2.0 voor de voorspelling van **status werkhervatting binnen 12 maanden in de klantgroep WW 55+**;  $n = 2339$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
28,7	95,0	0,50	78,7	123,7	218,7	0,82	0,14
<b>45,6</b>	<b>90,0</b>	<b>0,42</b>	<b>79,1</b>	<b>135,6</b>	<b>225,6</b>		
60,0	85,0	0,35	78,8	145,0	230,0		
67,3	80,0	0,31	76,9	147,3	227,3		
73,4	75,0	0,28	74,6	148,4	223,4		
79,3	70,0	0,24	72,3	149,3	219,3		
69,9	77,3	0,29	75,5	147,2	224,6		
75,0	73,0	0,26	73,5	148,0	221,0		
80,0	68,4	0,24	71,3	148,4	216,9		
85,0	61,6	0,21	67,3	146,6	208,2		
90,1	51,3	0,17	60,8	141,4	192,7		
95,0	36,3	0,13	50,7	131,2	167,5		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met het hoogste percentage correct voorspeld is donkerblauw.

Tabel 8.4b: Voorspellende eigenschappen van het klantgroep-specifieke model voor de voorspelling van **status werkhervatting binnen 12 maanden in de klantgroep WW 55+**;  $n = 2338$ <sup>1</sup>.

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
30,3	95,0	0,60	79,1	125,3	220,3	0,82	0,14
<b>49,0</b>	<b>90,0</b>	<b>0,45</b>	<b>79,9</b>	<b>139,1</b>	<b>229,1</b>		
58,1	85,0	0,37	78,4	143,1	228,1		
66,8	80,0	0,31	76,7	146,8	226,7		
73,4	75,0	0,26	74,6	148,4	223,4		
78,3	70,0	0,22	72,0	148,3	218,2		
69,9	78,2	0,29	76,2	148,1	226,4		
75,0	73,1	0,24	73,6	148,1	221,2		
80,0	68,5	0,20	71,3	148,5	217,0		
85,0	61,4	0,16	67,2	146,5	207,9		
90,1	54,3	0,13	63,1	144,4	198,8		
95,0	38,0	0,07	52,0	133,0	171,0		

<sup>1</sup> De persoon in de categorie vrouw getrouwd/woont samen, jongste kind thuis 0-6 jaar is verwijderd uit de analyses; niemand in de ontwikkelset en 1 persoon in de testset scoorde in deze categorie. AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met het hoogste percentage correct voorspeld is donkerblauw.

Voor de klantgroep WW 55+ zijn de voorspellende eigenschappen van Werkverkenner 2.0 (zie tabel 8.4a) vrijwel gelijk aan die van het *klantgroep-specifieke* model (zie tabel 8.4b). De globale maat (i.e. de AUC) van Werkverkenner 2.0 verschilt niet significant van de globale

maat van het klantspecifieke model ( $z = -1,35$ ;  $p = 0,18$ ). De Brierscores van beide modellen zijn gelijk; het percentage correct voorspeld is met Werkverkenner 2.0 iets lager (79,1%) dan met het klantgroep-specifieke model (79,9%). Deze percentages zijn hoger dan met de naïeve methode (75%). De maximale waarde van de som van de sensitiviteit en specificiteit is in het algemene model iets hoger (149,3) dan met het klantgroep-specifieke model (148,5).

Voor het beantwoorden van de vraag welke factoren belangrijk zijn voor deze klantgroep, inspecteren we het klantgroep-specifieke model, weergegeven in bijlage K, tabel K.2. Het specifieke model voor de klantgroep WW 55+ heeft in totaal 15 factoren. Dertien van deze 15 factoren komen ook voor in Werkverkenner 2.0 (zie tabel 8.3). Hoewel we een specifieke leeftijdsgroep hebben geselecteerd (55+), blijft de factor leeftijd belangrijk (i.e. op de 4<sup>e</sup> plaats). Nieuwe factoren voor deze klantgroep (t.o.v. de totale groep) zijn: Regio koplopervestiging en Financiële noodzaak om binnen 2 maanden werk te vinden. De relatieve bijdrage van ieder van deze nieuwe factoren is echter laag (i.e.  $\Delta R^2 < 0,01$ ; bijlage K, tabel K.2). We concluderen hieruit dat de toegevoegde waarde voor de diagnose van deze nieuwe factoren (t.o.v. de factoren van Werkverkenner 2.0) klein is.

### 8.2.3 Resultaten voor de klantgroep WW 35-min

Voor de klantgroep WIA 35-min zijn de voorspellende eigenschappen van Werkverkenner 2.0 (zie tabel 8.5a) vrijwel gelijk aan die van het *klantgroep-specifieke* model (zie tabel 8.5b). De globale maat (i.e. de AUC) van Werkverkenner 2.0 verschilt niet significant van de globale maat van het klantgroep-specifieke model ( $z = 0,18$ ;  $p = 0,86$ ). De Brierscore van Werkverkenner 2.0 is iets lager (d.w.z. gunstiger) dan het klantgroep-specifieke model (0,15 vs. 0,17); het percentage correct voorspeld met Werkverkenner 2.0 is iets hoger (78,8%) dan met het klantgroep-specifieke model (78,4%).

Tabel 8.5a: Voorspellende eigenschappen van het *algemene* model voor de voorspelling van **status werkhervatting binnen 12 maanden in de klantgroep WIA 35-min**;  $n = 292$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
18,0	94,8	0,50	78,8	112,8	207,6	0,72	0,15
31,2	90,0	0,39	77,7	121,2	211,2		
41,0	84,9	0,35	75,7	125,8	210,7		
49,2	80,1	0,31	73,6	129,3	209,4		
57,4	74,9	0,28	71,2	132,3	207,2		
60,7	70,1	0,24	68,2	130,8	200,9		
70,5	61,9	0,20	63,7	132,4	194,3		
75,4	58,4	0,18	62,0	133,9	192,3		
80,3	48,9	0,14	55,5	129,3	178,2		
85,3	41,6	0,11	50,7	126,8	168,4		
90,2	24,7	0,07	38,4	114,8	139,5		
95,1	18,6	0,05	34,6	113,7	132,3		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met het hoogste percentage correct voorspeld is donkerblauw.

Ook de maximale waarde van de som van de sensitiviteit en specificiteit is met Werkverkenner 2.0 iets hoger (133,9) dan met het klantgroep-specifieke model (132,4).

Het percentage correct voorspeld met de naïeve methode ligt het hoogst, namelijk op 79%. Er is echter nauwelijks verschil met het percentage correct voorspeld volgens Werkverkenner 2.0 (het verschil is 0,2%punt). Dit betekent dat we met toevoeging van de

factoren aan het model (of het nu het model van Werkverkenner 2.0 is of het klant-specifieke model) niet beter (maar ook niet slechter) voorspellen dan dat we iedereen aan de grootste categorie van status werkhervatting toewijzen (d.w.z. voor de klantgroep WIA 35-min: geen werkhervatting binnen 12 maanden).

Tabel 8.5b: Voorspellende eigenschappen van het klantgroep-specifieke model voor de voorspelling van **status werkhervatting binnen 12 maanden in de klantgroep WIA 35-min**;  $n = 292$ .

Sensitiviteit (SN)	Specificiteit (SP)	Afkapwaarde	% correct voorspeld	SN + SP	SN + 2SP	AUC	Brier
16,4	94,8	0,63	78,4	111,2	206,0	0,71	0,16
31,1	90,0	0,50	77,7	121,2	211,2		
37,7	84,8	0,40	75,0	122,6	207,4		
44,3	80,1	0,32	72,6	124,3	204,4		
55,7	74,9	0,25	70,9	130,6	205,5		
62,3	70,1	0,21	68,5	132,4	202,6		
70,5	56,7	0,14	59,6	127,2	183,9		
75,4	56,7	0,13	60,6	132,1	188,8		
80,3	51,1	0,11	57,2	131,4	182,5		
85,2	43,3	0,09	52,1	128,5	171,8		
90,2	34,2	0,06	45,9	124,4	158,6		
95,1	13,4	0,02	30,5	108,5	121,9		

AUC = Area-under-the-curve: onderscheidend vermogen van het model. De regel met het hoogste percentage correct voorspeld is donkerblauw.

Voor het beantwoorden van de vraag welke factoren belangrijk zijn voor deze klantgroep, inspecteren we het klantgroep-specifieke model, weergegeven in bijlage K, tabel K.3. Het specifieke model heeft in totaal 11 factoren. De belangrijkste 3 factoren van dit model komen ook voor als belangrijke factoren in het model van Werkverkenner 2.0: Maximale duur WW-recht, Samenstelling huishouden en Leeftijd (zie tabel 8.3). In totaal 8 van de 11 factoren in het klantgroep-specifieke model komen ook voor in Werkverkenner 2.0. De drie nieuwe factoren voor deze klantgroep zijn: Financiële noodzaak om binnen 2 maanden werk te vinden, Bemiddelingsberoep niveau en Belemmering: financiële problemen. De relatieve bijdrage van ieder van deze nieuwe factoren ligt tussen de 0,009 (Belemmering: financiële problemen) en 0,027 (Financiële noodzaak om binnen 2 maanden werk te vinden), zie bijlage K, tabel K.3.

Tot slot bekijken we opnieuw de volledige tabel 8.3 en concluderen dat een vijftal factoren belangrijk zijn voor zowel de totale groep als de drie klantgroepen: Leeftijd, Samenstelling huishouden, Maximale duur WW-recht, Werkzoekgedrag m.b.t. direct contact met werkgevers en Gezondheidsbeleving.

## 9 Referenties

- Abraham, C., & Michie, S. (2008). A taxonomy of behaviour change techniques used in interventions. *Health psychology, 27*(3), 379-387.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control, 19*(6), 716-723.
- Allison, P. (2010). Survival analysis. In G. R. Hancock, R. O. Mueller & L. M. Stapleton (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 413-424). New York (NY): Routledge.
- Ajzen, I., (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.
- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier life-styles? *Psychological Science in the Public Interest, 4*, 1-44.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.
- Blanson Henkemans, O. A. (2009). *ePartner for self-care: How to enhance eHealth with personal computer assistants* [Proefschrift]. Delft, Nederland: TU Delft.
- Blau, G. (1994). Testing a two-dimensional measure of job-search behavior. *Organizational Behavior and Human Decision Processes, 59*, 288-312.
- Bohner, G., & Wänke, M. (2002). *Attitudes and attitude change*. New York, NY: Taylor & Francis Inc.
- Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior, 23*(1), 463-477.
- Brouwer, S., Schellekens, J. M. H., Havinga, H., & van Brakel, K. (2011). *Voorspellers van Werkhervatting: een onderzoek onder werklozen in Noord-Holland*. Groningen/Amsterdam, Nederland: Rijksuniversiteit Groningen UMCG/UWV Kenniscentrum.
- Brouwer, S., Bakker, R. H., & Schellekens, J. M. H. (2015). Predictors for re-employment success in newly unemployed: A prospective cohort study. *Journal of Vocational Behavior, 89*, 32-38.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

De Groot, N., & Van der Klaauw, B. (2014). The effects of reducing the entitlement period to unemployment insurance benefits. *IZA Discussion Paper, No. 8336*, 1-44.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44, 837-845.

Dodou, D., & de Winter, J. C. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, 36, 487-495.

Eden, D., & Aviram, A. (1993). Self-efficacy training to speed reemployment: Helping people to help themselves. *Journal of Applied Psychology*, 78, 352-360.

Eggen, T. J. H. M., & Sanders, P. F. (1993). *Psychometrie in de praktijk*. Arnhem: Cito Instituut voor Toetsontwikkeling.

Evers, A., Lucassen, W., Meijer, R., & Sijsma, K. (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests (geheel herziene versie)*. Bekeken via [www.cotandocumentatie.nl](http://www.cotandocumentatie.nl).

Ford, M.E. (1992). *Motivating Humans: goals, emotions, and personal agency beliefs*. London: Sage.

Furnham, A., & Rawles, (1996). Job search strategies, attitudes to school and attributions about unemployment. *Journal of Adolescence*, 19, 355-369.

George, D., & Mallery, P. (2003). *IBM SPSS Statistics 21 step by step: A simple guide and reference*. Boston, MA: Allyn & Bacon.

Gurney, R. M. (1981). Leaving school, facing unemployment, and making attributions about the causes of unemployment. *Journal of Vocational Behavior*, 18, 79-91.

Hak, T., Van der Veer, K., & Jansen, H. (2004). *The Three-Step Test-Interview (TSTI): An observational instrument for pretesting self-completion questionnaires* [Erim Report Series-2004-029-ORG]. Rotterdam, Nederland: Erasmus Universiteit Rotterdam

He, J., & Van den Vijver, F. (2012). Bias and equivalence in cross-cultural research. Bias and Equivalence in Cross-Cultural Research. *Online Readings in Psychology and Culture*, 2(2). <http://dx.doi.org/10.9707/2307-0919.1111>.

Heagerty, P. J., & Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, 61(1), 92-105.

Henderson, R., & Keiding, N. (2005). Individual survival time prediction using statistical models. *Journal of Medical Ethics*, 31(12), 703-706.

Hermesen, S., & Renes, R. J. (2014). *Ontwerpen voor gedragsverandering*. Utrecht, Nederland: U Create, Centre of Expertise Creative Industry. Verkregen op 23 oktober 2016 van [https://issuu.com/coecreativeindustrie/docs/ontwerpen\\_voor\\_gedragsverandering\\_i](https://issuu.com/coecreativeindustrie/docs/ontwerpen_voor_gedragsverandering_i)

Hu, L., & Bentler, P. M. (2009). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.

Hyndman, R.J., & Fan, Y (1996). *The American Statistician*, 50(4), pp. 361-365.

Kanfer, R., Wanberg, C. R., & Kantrowitz, T. M. (2001). Job search and employment: A personality–motivational analysis and meta-analytic review. *Journal of Applied Psychology*, 86(5), 837-855.

Kline, R. B. (2004). *Principles and practice of structural equation modeling* (2<sup>nd</sup> ed.). New York, NY: Guilford.

Kopelman, R. E., Rovenpor, J. L., & Millsap, R.E. (1992). Rationale and Construct Validity Evidence for the Job Search Behavior Index: Because Intentions (and New Years Resolutions) Often Come to Naught. *Journal of Vocational Behavior*, 40, 269-287.

Kranenborg, K., Dusseldorp, E., & Sonke, C. (2017). *Rapportage gebruikersonderzoek UWV Werkverkenner* [TNO intern rapport R10293]. Leiden, Nederland: TNO.

Kuha, J. (2004). AIC and BIC comparisons of assumptions and performance. *Sociological Methods & Research*, 33 (2), 188-229.

Lagerveld, S. E., Brenninkmeijer, V., Blonk, R. W. B., Twisk, J., & Schaufeli, W. B. (2016). Predictive value of work-related self-efficacy change on RTW for employees with common mental disorders. *Occupational Environmental Medicine*, 0, 1-3, doi:10.1136/oemed-2016-104039.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Luteijn F., Starren J., & van Dijk H. (2000) *Handleiding Nederlandse Persoonlijkheds Vragenlijst, NPV (herziene uitgave)*. [Manual Dutch Personality Questionnaire, NPV (revised ed.)]. Lisse, Nederland: Swets Test Publishers.

MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 16-36). Thousand Oaks, CA: Sage.

Neerincx, M.A. & Lindenberg, J. (1999). Supporting individual situation awareness in webenvironments. *Ergonomie in uitvoering: de digitale mens. Congresboek Nederlandse Vereniging voor Ergonomie*, 1999, 144-155.

Neerincx, M.A., Lindenberg, J., & Rypkema, J.A. (1999). *Gebruiksvriendelijkheid van Internetdienst voor aangifte van loon- en omzetbelasting* [Rapport TM-99-C025]. Soesterberg, Nederland: TNO Technische Menskunde.

Neerincx, M.A., Ruijsendaal, M., & Wolff, M. (2001). Usability Engineering guide for Integrated Operation Support in Space Station Payloads. *International Journal of Cognitive Ergonomics*, 5 (3), 187-198.

Neter, J., M. H. Kutner, C. J. Nachtsheim, & Wasserman, W. (1996), *Applied linear statistical models* (4<sup>th</sup> ed.). Chicago, IL: Irwin.

NOA (2008). *Evaluatieonderzoek klantprofieltest* [Intern rapport]. Amsterdam: NOA.

Paulissen, R. T., Neerincx, M. A., Cremers, A. H. M., & Kranenburg, K. (2013). *Interaction Design Patterns for ePartners: Method and Example* [TNO rapport R10777]. Leiden, Nederland: TNO.

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL <https://www.R-project.org/>

Rogerson, P. A. (2001), *Statistical Methods for Geography*. London, UK: Sage.

Schellekens, J. M. H., Abma, F., Nijland, M., Postma, G., Van Bebber, J., Mulders, H., Havinga, H., & Van Brakel, K. (2007). *Voorspellers voor succesvolle werkhervatting. Een vergelijking tussen langdurig werklozen en snelle werkhervatters*. Groningen: Experimentele en Arbeidspsychologie RUG.

Schellekens, J. M. H., Langkamp, M., & De Vries, G. (2005). *Ontwikkeling competentievragenlijst voor uitzendkrachten: Tussenrapport voor ABU en UWV*. Groningen: Arbeidspsychologie RUG/Amsterdam: Kenniscentrum UWV.

Schwarz, G. E. (1978), Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461 – 464.

Sherer, M., Maddux, J. E., Mercadante, B., Prentice-Dunn, S. Jacobs, B., & Rogers, R.W. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports*, 51, 663-671.

Sijtsma, K. (2009). Over misverstanden rond Cronbachs alfa en de wenselijkheid van alternatieven. *De Psycholoog*, november, 561-567.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.

Tourangeau, R., Rips, L. C., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.

Van der Heul, H., Berendsen, E. M., Van der Eijken, B. C. J. L., & Vlek, R. J. J. (2003). *Op weg naar werk. Een evaluatie van de sluitende aanpak in de WW*. Amsterdam: UWV.

Van der Heul, H. (2006). Kennismemo 2006/11. *Vaststelling en evaluatie van de effectiviteit van reïntegratiemaatregelen voor WW cliënten*. Amsterdam: Kenniscentrum UWV.

Van Hooft, E. A. J. (2004). *Job seeking as planned behavior: In search of group differences*. Amsterdam: Vrije Universiteit.

Van Hooft, E. A. J., Born, M. P., Taris, T. W., Van der Flier, H., & Blonk, R. W. B. (2004). Predictors of job search behavior among employed and unemployed people. *Personnel*

*Psychology*, 57(1), 25-59.

Van Hooft, E. A. J., Born, M. P., Taris, T. W., Van der Flier, H., & Blonk, R. W. B. (2005). Bridging the gap between intentions and behavior: Implementation intentions, action control, and procrastination. *Journal of Vocational Behavior*, 66(2), 238-256.

Van Houwelingen, H. C., & Putter, H. (2011). *Dynamic prediction in clinical survival analysis*. Boca Raton, FL: CRC Press. ISBN 9781439835333.

Venables, W. N., & Ripley, B. D. (2002) *Modern Applied Statistics with S* (4<sup>th</sup> ed.). New York, NY: Springer.

Vinokur, A., & Caplan, R. D. (1987). Attitudes and social support: Determinants of job-seeking behavior and well-being among the unemployed. *Journal of Applied Psychology*, 17, 1007-1024.

Vroom, V. H. (1964). *Work and motivation*. New York, NY: John Wiley & Sons.

Wanberg, C. R., Song, Z, & Hough, L. M. (2002). Predictive validity of a multidisciplinary model of reemployment success. *Journal of Applied Psychology*, 87, 1100-1120.

Weiner, B. (1980). *Human Motivation*. New York, NY: Holt, Rinehart & Winston.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York, NY: Springer-Verlag.

Weinstein, M. C., & Fineberg, H. V. (1980). *Clinical decision analysis*. Philadelphia, PA: W.B. Saunders.



