

WSE Report

# Het statistisch voorspellen van het risico op langdurige werkloosheid

Vicky Heylen

November 2011

## Inhoudstafel

### Het statistisch voorspellen van het risico op langdurige werkloosheid

Abstract	3
1. Inleiding	4
2. Diagnose-instrumenten	6
3. Profiling in het buitenland	7
4. Data en methodologie	8
5. De (theoretische) kans op langdurige werkloosheid	13
6. Conclusie	19
Bibliografie	22

### Lijst van tabellen

Tabel 1	Werkloosheidsgraad voor 15-64-jarigen (jaarlijks gemiddelde)	4
Tabel 2	Aandeel langdurig werklozen in de totale werklozenbevolking	5
Tabel 3	Duur van de werkloosheidsperiodes die van start gingen in 2007	9
Tabel 4	Persoonskenmerken naar werkloosheidsduur	12
Tabel 5	Logit kans op langdurige werkloosheid bij instroom (marginale effecten, standaard fout, z-waarde)	14
Tabel 6	Voorspelvermogen van het logit model geschat bij instroom in de werkloosheid	17
Tabel 7	Logit kans op langdurige werkloosheid bij mannen na 3 en 6 maanden werkloosheid (marginale effecten, standaard fout, z-waarde)	18
Tabel 8	Voorspelwaarde van het model bij instroom, na 3 en 6 maanden	19

### Lijst van grafieken

Grafiek 1	Overlevingscurves voor de uitstroom uit NWWZ en de uitstroom naar werk	10
Grafiek 2	Empirische uitstroomkans uit NWWZ en naar werk	11
Grafiek 3	Voorspelde kans op langdurige werkloosheid	16

## Het statistisch voorspellen van het risico op langdurige werkloosheid

### **Abstract:**

In de context van een recessie met een stijgende werkloosheid en krimpende budgetten wordt het nog belangrijker activeringsmiddelen in te zetten op de meest efficiënte manier. Het snel en accuraat diagnosticeren van werklozen kan hierbij een belangrijke rol spelen. In deze paper wordt een statistisch voorspellingsmodel om de kans op langdurige werkloosheid te schatten ontwikkeld. Een cohorte instromers uit 2007 wordt in de dataset van de VDAB geïsoleerd en opgevolgd tot september 2010. Factoren zoals geslacht, leeftijd, opleidingsniveau, woonplaats, etniciteit, arbeidshandicap, beroeps- en werkloosheidsgeschiedenis hebben alle een significante invloed op het werkloos blijven gedurende 12 maanden en meer. De voorspelkracht van het model blijkt echter eerder beperkt. Wanneer ieder met een kans hoger dan 20% als risicowerkloze beschouwd wordt, wordt slechts 23.1% van de mannelijke langdurig werklozen en 30.9% van de vrouwelijke langdurig werklozen als dusdanig geclassificeerd. De voorspelkracht van het model kan verder verhoogd worden door meer relevante kenmerken te identificeren en te meten, of door de statistische diagnosticering uit te stellen naar een later moment in de werkloosheidsperiode.

## 1. Inleiding

In de context van een recessie met steeds stijgende werkloosheidscijfers en krimpende budgetten wordt een efficiënte inzet van deze middelen steeds belangrijker. Arbeidsbemiddelingsdiensten trachten daarom zo snel mogelijk de personen met de hoogste kans om langdurig afhankelijk te zijn van uitkeringen te identificeren. Een manier om zulke inschatting te maken is door statistisch profileren of diagnosticeren. Bij het statistisch profileren gaat men op basis van een aantal kenmerken van werklozen en de lokale arbeidsmarkt een theoretisch risico op langdurige werkloosheid bepalen. Aan deze score kan dan de deelname aan activeringsmaatregelen gekoppeld worden.

Vooraf in periodes van recessie is het vermijden van (langdurige) werkloosheid een onderwerp dat hoog op de agenda staat. Men vreest immers voor het zogenaamde hysteresiseffect waarbij een conjunctuurgebonden toename in de werkloosheid evolueert naar een structurele toename van de werkloosheidsgraad. Dit heeft te maken met het feit dat de kans om uit te stromen uit de werkloosheid kleiner wordt naarmate iemand langer werkloos is. Het aandeel werklozen in de Belgische beroepsactieve bevolking liep sinds de start van de crisis in 2008 met 1.4 pptn. op tot 8.4% in 2010 (cfr. tabel 1). In Vlaanderen bedroeg deze stijging 1.3 pptn. In vergelijking met de gemiddelde stijging in E.U. 27 met 2.6 pptn. viel deze stijging nog mee. Dit wordt doorgaans verklaard door het gebruik van het stelsel van tijdelijke werkloosheid en andere anti-crisismaatregelen (vb. crisistijds-krediet, tijdelijke crisisaanpassing van de arbeidsduur, enz.) (BNB 2011). Duitsland is vrijwel het enige Europese land waar de werkloosheidsgraad stabiel bleef en in 2010 met 7.2% zelfs lager ligt dan in 2008 toen de werkloosheidsgraad 7.6% bedroeg.

Tabel 1 Werkloosheidsgraad voor 15-64-jarigen (jaarlijks gemiddelde)

	2007	2008	2009	2010
België	7.5	7.0	8.0	8.4
Vlaams Gewest	4.4	3.9	5.0	5.2
Waals Gewest	10.5	10.1	11.2	11.5
Brussels H. Gewest	17.2	16.0	15.9	17.4
E.U. 27	7.2	7.1	9.0	9.7
Duitsland	8.7	7.6	7.8	7.2
Frankrijk	8.0	7.4	9.1	9.3
Nederland	3.2	2.7	3.4	4.5
Verenigd Koninkrijk	5.4	5.7	7.7	7.9

Bron: FOD Economie - ADSEI EAK, Eurostat LFS (bewerking Steunpunt WSE/Departement WSE)

De evolutie van de langdurige werkloosheid wordt algemeen gezien als een eerste indicatie van hysteresis. Doorgaans wordt iemand gedefinieerd als langdurig werkloos wanneer die langer dan een jaar werkloos was. Tabel 2 geeft de evolutie van het aandeel langdurig werklozen binnen de totale werklozenpopulatie.

Tabel 2 Aandeel langdurig werklozen in de totale werklozenbevolking

	2007	2008	2009	2010
België	50.4	47.6	44.2	48.8
Vlaams Gewest	37.5	36.7	30.4	37.0
Waals Gewest	57.0	51.7	52.2	55.5
Brussels H. Gewest	57.6	55.6	52.6	56.2
EU-27	42.7	37.0	33.2	39.9
Duitsland	56.6	52.6	45.5	47.3
Frankrijk	40.2	37.6	35.2	40.1
Nederland	39.4	34.8	24.8	27.6 <sup>1</sup>
Verenigd Koninkrijk	23.8	24.1	24.5	32.7

<sup>1</sup> Tussen 2009 en 2010 is er een tijdreeksbreuk voor de Nederlandse gegevens.

Bron: Eurostat LFS (Bewerking Steunpunt WSE/Departement WSE)

Het aandeel langdurig werklozen daalde in eerste instantie in 2009. Dit heeft natuurlijk te maken met de verhoogde instroom in de werkloosheid, zoals tabel 1 reeds aangaf. Vanaf 2010 zien we voor alle landen een stijging. Blijkbaar hebben de personen die werkloos werden of intraden in de arbeidsmarkt tijdens de recessie het zeer moeilijk om werk te vinden en stijgt het aantal langdurig werklozen meer dan proportioneel. België bevindt zich tussen de koplopers van Europa, bijna de helft van de werklozen langdurig werkloos. Er blijken wel grote verschillen te zijn tussen de gewesten. In het Vlaams gewest bedroeg het aandeel langdurige werklozen in de totale werklozen bevolking 37.0%, in het Waals gewest was dit 55.5% en in het Brussels Hoofdstedelijk Gewest zelfs 56.2%.

Om de werkloosheid, en dan vooral de langdurige werkloosheid terug te dringen werden allerlei activerende maatregelen in het leven geroepen. Budgettair is het echter niet mogelijk alle instromers in de werkloosheid een activerende behandeling te geven. Om de effectiviteit van deze maatregelen te verhogen, is het van belang dat men de juiste mensen op het juiste moment bereikt. Dit kan men nastreven door te differentiëren tussen de cliënten. In vele ons omringende landen ziet men statistisch profileren als een manier om zulke differentiatie te bewerkstelligen. In deze paper gaan we na hoe de kans op langdurige werkloosheid van de Vlaamse werklozen kan voorspeld worden aan de hand van de gegevens die geregistreerd worden wanneer ze zich bij het begin van hun werkloosheidsperiode aanmelden als werkzoekende bij de VDAB (Vlaamse Dienst voor Arbeidsbemiddeling).

Deze paper is gestructureerd als volgt. De tweede paragraaf beschrijft de verschillende diagnose instrumenten, waarvan statistische profiling er één is. De derde paragraaf geeft een overzicht van de landen die potentieel langdurig werklozen trachten te identificeren aan de hand van statistische profiling. We geven hier, indien mogelijk, ook aan hoe deze diagnosemethode geëvalueerd wordt. Een beschrijving van de gebruikte data en methodologie volgen in paragraaf 4. In paragraaf 5 tonen we dan de resultaten van de logistische regressie, die de theoretische kans op langdurig werkloosheid voor Vlaamse werkzoekenden, berekend. Hierbij wordt verder aandacht besteed aan de voorspelwaarde van dit model. De zesde paragraaf sluit deze paper af met een samenvatting van de belangrijkste conclusies.

## 2. Diagnose-instrumenten

Hasluck (2008) onderscheidt vier verschillende diagnose-instrumenten die gebruikt worden om de inzetbaarheid van werkzoekenden op de arbeidsmarkt in te schatten en/of activerende maatregelen toe te kennen:

- *Door regels*: vooraf worden regels opgesteld die de cliënten op basis van hun karakteristieken indelen in bepaalde groepen. Elk van deze groepen, bijvoorbeeld de laaggeschoolde jongeren of de allochtone vrouwen, kunnen dan in verschillende mate toegang krijgen tot activeringsmaatregelen.
- *Door de consulent*: de consulent beslist volledig autonoom aan welke diensten een cliënt toegewezen wordt.
- *Door screening*: de consulent gebruikt testen om de inzetbaarheid van de cliënt in te schatten.
- *Door statistische profiling*: op basis van een statistische regressie tracht men de werkloosheidsduur of de kans op langdurige werkloosheid te voorspellen. Op basis van deze voorspelling deelt men de cliënten dan in in verschillende groepen.

In dit rapport zoomen we in op deze laatste vorm van diagnosticeren. Bij statistische profiling tracht men aan de hand van een regressiemodel en op basis van de kenmerken van de werkzoekende een voorspelling te maken over hun kans op langdurige werkloosheid. Een voldoende voorspelkracht van zulk model, dit is de nauwkeurigheid waarmee deze voorspellingen kunnen gedaan worden, is natuurlijk essentieel voor dit diagnose-instrument.

Idealiter moet rekening gehouden worden met *alle* factoren die de kans op het vinden van werk bepalen. Ongetwijfeld zou zulke voorspelling dus niet alleen gebaseerd moeten zijn op vaak gemakkelijk kwantificeerbare gegevens zoals leeftijd, geslacht, opleiding, werkervaring, enz., maar ook op moeilijker waarneembare en/of meetbare factoren als motivatie, zelfvertrouwen, netwerk, enz. Het probleem is natuurlijk dat niet alle relevante factoren kunnen geobserveerd worden in de reële wereld en (administratieve) datasets.

Aangezien het dus een utopie is te denken dat bij instroom in de werkloosheid voor elk individu de kans op werkloosheid perfect voorspeld kan worden, wordt statistisch profileren gebruikt om de werklozen te identificeren die een sterk verhoogd risico hebben om langdurig werkloos te worden. Eén van de belangrijkste voordelen van dit diagnose-instrument is dat het om een objectieve maatstaf gaat die voor elk individu op dezelfde wijze kan toegepast worden. Verder geeft een statistische profiling snel en goedkoop een eenduidig antwoord op de vraag 'heeft deze werkzoekende een hoge kans om langdurig werkloos te worden?'. Het is dus mogelijk het risico op langdurige werkloosheid voor grote aantallen werkzoekende in te schatten. Vanzelfsprekend zijn er ook enige nadelen verbonden aan deze manier van diagnosticeren. Ten eerste is er de vraag naar de accuraatheid en precisie: hoe juist en nauwkeurig zijn de voorspellingen in het model. Ten tweede is een statistische diagnose goedkoop in gebruik, doch de ontwikkelingskosten (vb. het specificeren van het juiste model, het meten van de relevante variabelen) en de opstartkosten (vb. het opleiden van de consulenten) kunnen erg hoog zijn. Daarenboven moet ook in het achterhoofd gehouden worden dat een model ontwikkeld voor een bepaalde regio in een bepaalde fase van de conjunctuur, mogelijk zeer slecht presteert in een andere regio of een andere fase van de conjunctuur. In die zin moet men ook rekening houden met terugkerende onderhoudskosten.

Algemeen wordt aan de hand van een voorspellingsmodel dus ingeschat wat de kans, tussen 0 en 100%, is dat een werkloze langdurig werkloos zal worden. Vervolgens moeten beleidsmakers zich buigen over de vraag buigen wanneer bijkomende acties gerechtvaardigd zijn. Is dit wanneer de kans op langdurige werkloosheid 50% geschat wordt, of reeds wanneer deze kans 30% overschrijdt. Vanzelfsprekend hangt deze beslissing samen met beleidsdoelstellingen en budgetten.

### 3. Profiling in het buitenland

De Verenigde Staten worden vaak beschreven als pioniers in het (statistisch) diagnosticeren. Sinds 1993 is elke staat immers verplicht om bij het begin van de werkloosheidsperiode in te schatten welke personen hun werkloosheidsuitkeringen, die beperkt zijn tot 26 weken, tot het einde zullen opnemen. De meeste staten gebruiken hiervoor statistische profiling en schatten een regressiemodel dat de kans om meer dan 26 weken werkloos te blijven modelleert. Als verklarende variabelen gebruikt men lokale economische indicatoren en enkele persoonsgegevens (opleidingsniveau, loopbaan, sector van laatste tewerkstelling, enz.<sup>1</sup>). Enkel op basis van deze inschatting worden de werklozen verplicht zich al dan niet aan te melden bij een bepaalde arbeidsbemiddelingsdienst of opleidingscentrum. De consulenten op deze diensten beslissen dan wel over welk activeringstraject de werklozen kan volgen. In die zin vertrouwt men dan ook maar ten dele op de statistische diagnose (Bimrose et al. 2007, Debauche & Georges 2007). Evaluaties geven aan dat de voorspellende waarde van de modellen hoog is en dat de statistische diagnose ertoe bijdraagt dat de werkloosheidsduur gevoelig verlaagd wordt (Black et al. 2003, O'Leary & Eberts 2010).

Het huidige Australische diagnoseinstrument, Job Seeker Classification Instrument (JSCI), werd ontwikkeld in 1998 en sindsdien geregeld geëvalueerd en geüpdatet. Bij de registratie van de werkloze wordt dit instrument ingezet om de personen met de grootste kans op langdurige werkloosheid te identificeren. De gebruikte verklarende variabelen in dit logistische model zijn leeftijd, geslacht, opleidingsniveau, taal en geletterdheid, recente werkervaringen, woonplaats, handicap, huishouden en mobiliteit eventueel aangevuld met aantal risicofactoren (lage eigenwaarde, lage motivatie, onverzorgd voorkomen, enz.) die een persoonlijke inschatting van de consulent vereisen<sup>2</sup>. Australië is daarmee het enige land waar men zogenaamde zachte kenmerken meeneemt in de analyse. Aan de hand van de score die een werkloze behaalde, wordt een gepaste vorm van arbeidsbemiddeling aangeboden. Degenen met slechts een beperkt risico op langdurige werkloosheid krijgen vacatures toegezonden. Degenen met een gemiddeld risico wordt een opleiding of training aangeboden. Voor degenen met een hoog risico wordt een activeringstraject met intensieve begeleiding opgestart. De budgetten die toegewezen worden aan de verschillende actoren zijn ook gedifferentieerd naargelang de gemiddelde score van het doelpubliek.

In Denemarken ging men in 2004 van start met een statistisch diagnose-instrument (Rosholm et al, 2004). In tegenstelling tot de modellen gebruikt in de VS en Australië ging men hier niet uit van een voorspellingsmodel dat aangeeft of een bepaalde werkloosheidsduur overschreden zal worden, maar van een duurmodel, dat tracht te voorspellen of men nog werkloos zal zijn binnen zes maanden, gegeven de reeds verstreken werkloosheidsduur. Verklarende variabelen die gebruikt werden bij deze schatting zijn leeftijd, geslacht, woonplaats, burgerlijke staat, gezondheid, opleidingsniveau, etniciteit, werkervaring en voorgeschiedenis in de werkloosheid. Het systeem werd echter niet geaccepteerd door de consulenten en werd dan ook afgeschaft. Momenteel wordt enkel de mening van de consulent meegenomen in de diagnose (Bimrose et al. 2007).

Naar aanleiding van de Harz-hervormingen werd in 2005 in Duitsland van start gegaan met het statistisch profileren van werklozen om de interactie tussen consulenten en cliënten te uniformiseren. Al bij het eerste gesprek, dat tot drie maanden voor de uiteindelijke instroom in de werkloosheid kan plaatsvinden, wordt de afstand tot de arbeidsmarkt gemeten. Hiervoor wordt beroep ge-

<sup>1</sup> Wettelijk mogen leeftijd, geslacht en etniciteit niet opgenomen worden in het voorspellingsmodel.

<sup>2</sup> [www.deewr.gov.au](http://www.deewr.gov.au)

daan op een aantal testen<sup>3</sup>. Initieel werden de werklozen ingedeeld in vier categorieën op basis van hun ingeschatte afstand tot de arbeidsmarkt, vanaf 2009 werd dit uitgebreid naar zes. De uitslag van het statistische programma kan hierbij steeds aangepast worden door de consulent. Dit gebeurt doorgaans wanneer er sprake is van bepaalde problemen (psychologische problemen, problemen van alcohol- of drugsmisbruik) die omwille van privacyredenen niet mogen geregistreerd worden in administratieve databanken (Mosley 2010). Aan de hand van de categorie waartoe men behoorde, werden de gepaste activeringsmaatregelen en de frequentie van de contacten met de consulent bepaald. Duitsland is één van de weinige landen waar al deze stappen verenigd zitten in één softwareprogramma, VerBIS.

In Frankrijk wordt statistisch profileren gebruikt sinds 2006. Om het risico op langdurige werkloosheid in te schatten wordt de uitkomst van de beslissingsboom evenwel steeds gecombineerd met de expertise van de consulent. Om het theoretisch risico op langdurige werkloosheid te schatten maakt men gebruik van de volgende achtergrondkenmerken: leeftijd, geslacht, burgerlijke staat, nationaliteit, arbeidsverleden, werkloosheidsgeschiedenis, voorkeursjob, recht op uitkering en lokale arbeidsmarktsituatie. Op basis van deze informatie worden de werkzoekenden ingedeeld in drie grote groepen; degenen met een laag, gemiddeld en hoog risico op langdurige werkloosheid. Vervolgens volgt een gesprek met de consulent van de arbeidsbemiddelingsdienst waarbij gepeild wordt naar eerder zachte kenmerken (o.a. gezondheid en motivatie) om zo de diagnose te verfijnen of eventueel bij te sturen. Uiteindelijk besluit de consulent dan ook welk traject aangeboden wordt: een snel zoektraject, een activeringstraject, een intensief begeleidingstraject, een traject naar ondernemerschap of een intensief traject naar de arbeidsmarkt (Debauche & Georges, 2007).

In Finland gebruikt men vanaf 2007 een statistisch diagnose-instrument om de kans op langdurige werkloosheid te bepalen. De verklarende variabelen die men opneemt zijn: leeftijd, woonplaats, opleidingsniveau, arbeidshandicap, werkloosheidsgeschiedenis, vorig beroep en de reden van de beëindiging ervan. De uitkomst van het model verplicht de consulent of werkloze echter tot niets. Bovendien staan de consulenten vrij sceptisch tegenover de accuraatheid van de voorspelling. In de praktijk heeft het statistisch diagnose-instrument dan ook slechts een zeer beperkte invloed (Riipinen, 2011).

Recent werden, tot slot, pilootprojecten opgestart in Ierland en Zweden. In Ierland werden initieel alle werklozen die meer dan drie maanden werkloos waren doorverwezen naar FÁS voor een begeleidingstraject. Aangezien evaluatiestudies (O'Connell et.al. 2010) echter aangaven dat men zowel op het vlak van efficiëntie als billijkheid veel kon gewonnen worden door de werklozen met de hoogste kans op langdurige werkloosheid aan het begin van de werkloosheidsperiode te identificeren, gebruikt men sinds 2011 een statistisch model om zulke kans te voorspellen. In Zweden gaat men op basis van administratieve gegevens de kans op langdurige werkloosheid, zijnde meer dan zes maanden werkloos, berekenen. De mening van de consulent wordt wel meegenomen in deze schatting. Zo deelt men de werklozenpopulatie in in vier groepen die elk specifieke activeringsmaatregelen toegewezen krijgen.

#### 4. Data en methodologie

Voor deze studie maken we gebruik van de gegevens die de VDAB van elke werkzoekende verzamelt in de CVS (Clientvolgsysteem)-databank. Hiervoor werd een cohorte die werkloos werd in

---

<sup>3</sup> Momenteel loopt een pilootproject KodiaK waarbij men de zachte kenmerken van de werkzoekenden ook meer en meer in kaart tracht te brengen. Vanaf 2012 zou deze module algemeen geïntegreerd worden in VerBIS.



2007 afgebakend<sup>4</sup>. De aldus afgebakende steekproef bevat 210 658 werkloosheidsperiodes. Aangezien sommige personen in 2007 meer dan eens instroomden in de werkloosheid, hebben deze periodes betrekking op 170 428 unieke werkzoekenden. Deze personen werden dan tot september 2010 geobserveerd in deze databank. Zo wordt duidelijk welk van deze personen uitstroomden en eventueel terug instroomden in de werkloosheid in de periode januari 2007 tot september 2010. Door de koppeling van de VDAB-dataset aan de DIMONA (déclaration immédiate–onmiddellijke aangifte)-gegevens is daarenboven geweten of de werkzoekende daadwerkelijk uitstroomt naar werk. Samenvattend kunnen we stellen dat van de 210 658 werkloosheidsperiodes, 170 420 (of 80.9%) eindigt met een uitstroom naar werk, 35 253 (of 16.7%) met een uitstroom naar niet-werk (zorgtaken, studies, buitenland, (vervroegd) pensioen, enz.) en 4 985 (of 2.4%) periodes nog steeds lopen op het einde van onze observatieperiode.

De CVS-databank bevat voornamelijk informatie over de zogenaamde harde kenmerken van de werkzoekende: het geslacht, de leeftijd, de woonplaats, het hoogst behaalde diploma, de talenkennis, de mobiliteit, de etniciteit en eventuele beroepservaring. Bovendien kunnen we ook nagaan of de persoon reeds voorkwam in de CVS-databank vóór 2007 en zo een werkloosheidsgeschiedenis samenstellen.

Iemand wordt beschouwd als langdurig werkloos wanneer die meer dan twaalf maanden onafgebroken geregistreerd was als niet werkende werkzoekende. Hierin verschilt onze definitie van die van de VDAB die onderbrekingen tot drie maand toelaten, alvorens de begindatum van de werkloosheidsperiode te resetten. Wanneer iemand dus een maand aan het werk gaat en nadien terug werkloos wordt, begint in onze benadering een nieuwe werkloosheidsperiode, terwijl voor de VDAB de oude werkloosheidsperiode gewoon terug opgenomen wordt. Zeer langdurig werklozen zijn meer dan 24 maanden onafgebroken niet werkend, werkzoekend. Tabel 3 geeft een overzicht van de duur van de werkloosheidsperiodes in de gebruikte steekproef

Tabel 3 Duur van de werkloosheidsperiodes die van start gingen in 2007

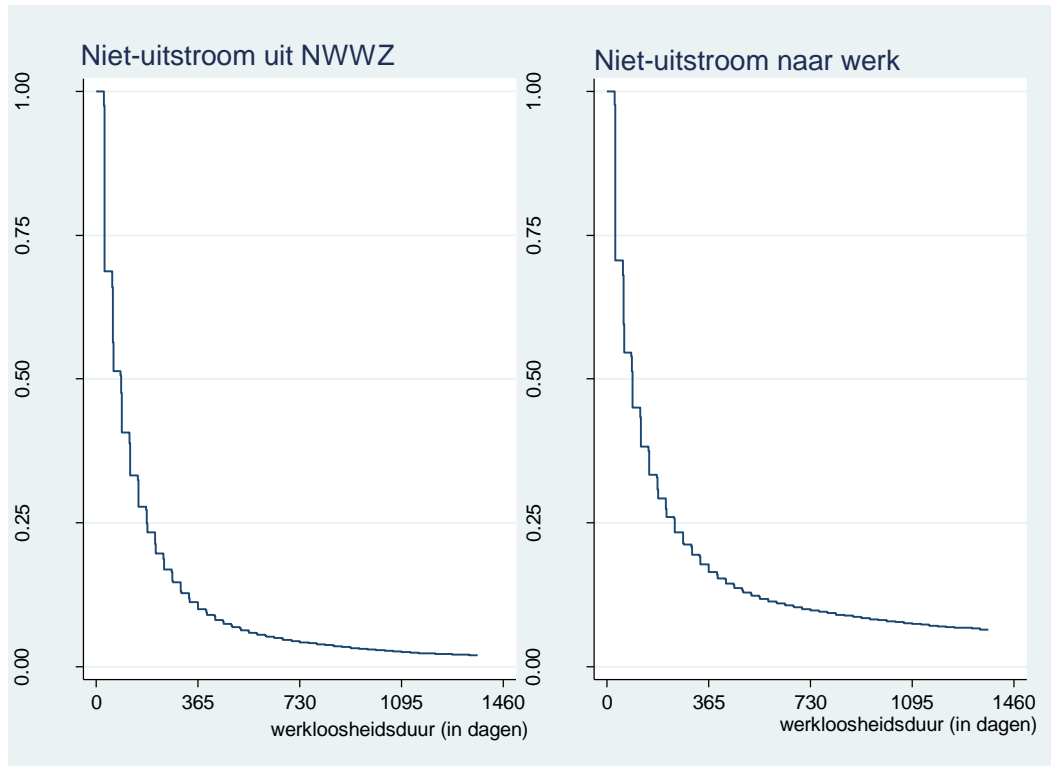
Werkloosheidsduur	Aantal	%
Uitstroom uit NWWZ		
< 1 jaar wl	187509	89.0
≥ 1 jaar wl	23149	11.0
1-2 jaar wl	13732	6.5
≥ 2 jaar wl	9417	4.5
Totaal	210658	100.0
Uitstroom naar werk		
< 1 jaar wl	159314	93.5
≥ 1 jaar wl	11106	6.5
1-2 jaar wl	8712	5.1
≥ 2 jaar wl	2394	1.4
Totaal	170420	100.0

Bron: VDAB en eigen bewerkingen HIVA

<sup>4</sup> We selecteerden enkel de werkloosheidsperiodes die starten met de werkloosheidscategorie '00', i.e. een werkloze die een werkloosheidsuitkering ontvangt. Daarnaast wordt als voorwaarde opgelegd dat de persoon in de maand voorafgaand aan de start van de werkloosheidsperiode, hetzij aan het werk was, hetzij nog niet gekend was door de VDAB.

We vinden dat 11% van de werkloosheidsperiodes die startten in 2007 een jaar later nog steeds lopen. 13 732 van deze periodes worden afgesloten voor de 2-jaar grens, 9 417 dueren ook na twee jaar verder. Onder de 170 420 werkloosheidsperiodes die eindigen door het vinden van een baan, vinden we merkkelijk minder langdurige werkloosheidsperiodes, namelijk 6.5%, terug.

In grafiek 1, die de overlevingsfunctie voor de werkloosheidsperiodes die van start gingen in 2007 uitzet, worden de werkloosheidsduur op een alternatieve manier in beeld gebracht. De overlevingsfunctie geeft voor elke werkloosheidsduur aan welk percentage van de uitstroom nog niet uitstroomde of nog niet uitstroomde naar werk.

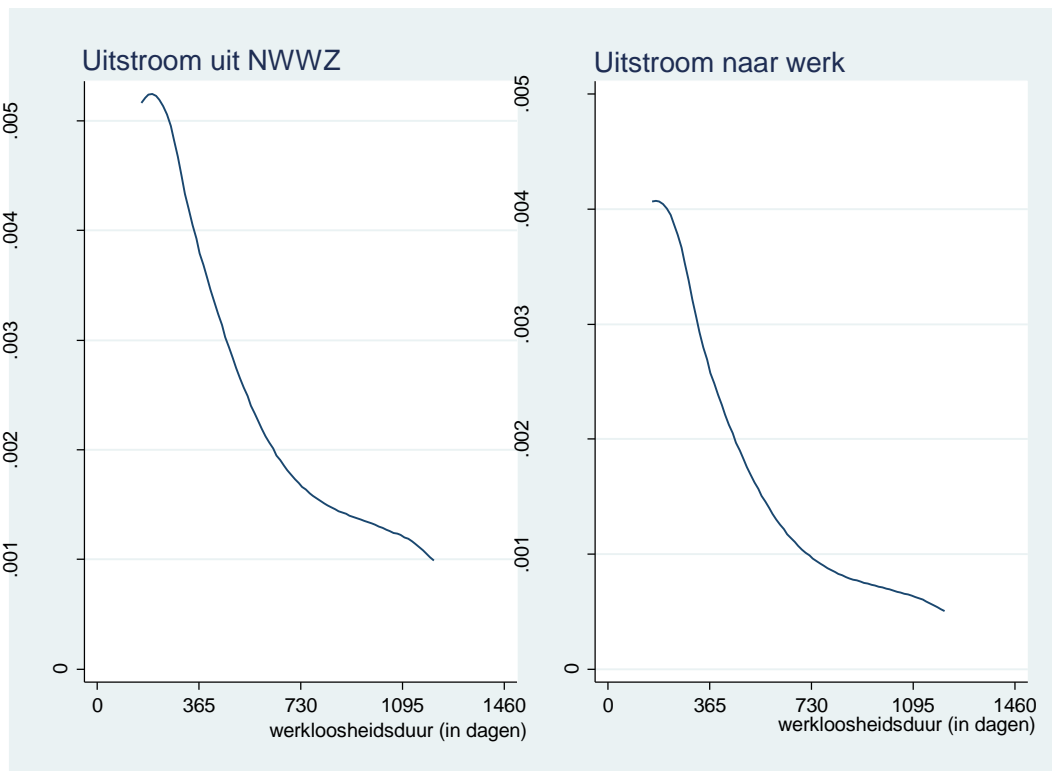


Bron: VDAB en eigen bewerkingen HIVA

Grafiek 1 Overlevingscurves voor de uitstroom uit NWWZ en de uitstroom naar werk

In het begin van de werkloosheidsperiode is iedereen nog werkloos. Dit aandeel zakt echter snel; van de cohorte instromers stroomt een groot deel relatief snel uit. De 'trapjes' worden kleiner naarmate de werkloosheidsduur toeneemt; dit geeft aan dat steeds minder werkzoekenden uitstromen.

Deze evolutie kan ook weergegeven worden door het uittekenen van de empirische uitstroomkans op een bepaald moment in een werkloosheidsperiode. Deze grafieken worden weergegeven in grafiek 2.



Bron: VDAB en eigen bewerkingen HIVA  
 Grafiek 2 Empirische uitstroomkans uit NWWZ en naar werk

Ook deze curve is een duidelijke weergave van de dalende uitstroomkans, bij een oplopende werkloosheidsduur. Na twee jaar is de uitstroomkans naar werk, bijvoorbeeld, nog maar een vierde van de uitstroomkans na een half jaar. Let wel, het gaat hier om een geobserveerde uitstroomkans, die geen conclusies toelaat naar de reden van deze daling. Het zou kunnen dat de uitstroomkans daalt ten gevolge van de oplopende werkloosheidsduur, maar eerder onderzoek (Heylen 2010) op dezelfde data geeft aan dat hier veeleer een selectie-effect speelt. Bepaalde werkzoekenden hebben bij de start van hun werkloosheidsperiode reeds een kleinere kans om werk te vinden, naarmate de werkloosheidsduur stijgt, zal deze groep, de facto, meer en meer vertegenwoordigd zijn binnen een bepaalde cohorte en zal de gemiddelde uitstroomkans voor de groep steeds lager liggen.

In tabel 4 wordt alvast een overzicht gegeven van een aantal persoonskenmerken die aan het begin van de werkloosheidsperiode voor elke werkzoekende geregistreerd worden. We maken hierbij een onderscheid tussen zij die uitstromen vóór ze 12 maanden werkloos zijn en zij die langer werkloos blijven.

Tabel 4 Persoonskenmerken naar werkloosheidsduur

	< 1 jaar wl	≥ 1 jaar wl
Aantal	187509	23149
Geslacht		
Man	51.6	49.4
Vrouw	48.4	50.6
Leeftijd		
< 25	25.4	10.3
25-49	67.2	69.2
≥ 50	7.4	20.5
Opleidingsniveau		
Laaggeschoold	44.7	50.2
Middengeschoold	36.7	35.7
Hooggeschoold	18.6	14.1
Etniciteit		
Autochtoon	91.3	87.2
Allochtoon (E.U.)	4.0	5.0
Allochtoon (niet E.U.)	4.7	7.8
Kennis Nederlands		
Ja	97.3	94.8
Neen	2.7	5.2
Arbeidsgehandicapt		
Ja	6.4	10.5
Neen	93.6	89.5
Woonplaats		
Grootstedelijk	34.5	39.3
Sterk functioneel verstedelijkt	11.4	10.3
Matig functioneel verstedelijkt	20.4	18.1
Zwak functioneel verstedelijkt	33.8	32.3
Rijbewijs B		
Ja	74.0	69.5
Neen	26.0	30.6
Auto ter beschikking		
Ja	54.1	45.1
Neen	45.9	54.9
Aantal maanden werkloos in 24mnd voor instroom		
0	24.7	38.4
1 tot 6	30.0	17.9
7 tot 12	23.6	18.1
13 tot 24	21.7	25.6
Beroepservaring in voorkeursberoep		
Geen	31.1	33.0
< 2 jaar	30.2	23.0
2 tot 5 jaar	17.0	13.6
> 5 jaar	21.7	30.4

Bron: VDAB en eigen bewerkingen HIVA

Op basis van deze bivariate analyse zien we reeds duidelijke verschillen in persoonskenmerken naargelang men al dan niet tot de groep langdurige werklozen behoort. Met betrekking tot geslacht blijven de verschillen vrij beperkt; iets meer vrouwen dan mannen worden langdurig werkloos. Voor

leeftijd maken we een onderscheid tussen personen jonger dan 25, personen tussen 25 en 50 en personen van 50 of ouder. Het is vooral deze laatste groep die opvallend meer voorkomt in de groep van werkzoekend die reeds meer dan één jaar werkloos, met name 20.5% t.o.v. 7.4% bij de kortdurig werklozen. Verder zien we een duidelijke oververtegenwoordiging van het aantal laaggeschoolden (i.e. de personen wier hoogst behaalde diploma maximaal lager secundair onderwijs is); zij maken de helft van de groep langdurig werklozen uit. Allochtonen, en dan vooral allochtonen van buiten de E.U., komen relatief gezien ook meer voor bij de groep die minstens een jaar werkloos is. De groep personen die geen Nederlands spreken is met 5.2% in verhouding bijna dubbel zo groot onder de langdurige dan onder de kortdurig werklozen. Ook vinden we veel meer arbeidsgehandicapten terug bij de langdurig werklozen; één op de tien langdurig werklozen wordt beschouwd als arbeidsgehandicapt. In 2007 werd arbeidshandicap in de CVS-data gedefinieerd als ofwel BuSO (Buitengewoon Secundair Onderwijs) gevolgd, ofwel een VAPH (Vlaams Agentschap voor Personen met een Handicap)-nummer ofwel als beperkt of zeer beperkt arbeidsgeschikt gelabeld door de VDAB-consulent. Voor de woonplaats van de werkloze wordt bekeken of deze kan omschreven worden als grootstedelijk, sterk, matig of zwak functioneel verstedelijkt. Om de graad van functionele verstedelijking te bepalen gaat men na wat de commerciële-, school- en werkfunctie van een bepaalde gemeente is. Uit tabel 4 komt duidelijk naar voor dat inwoners van grote steden oververtegenwoordigd zijn in de langdurige werkloosheid. Een beperkte mobiliteit, in de dataset gemeten door het ontbreken van een rijbewijs en/of een auto om te gaan werken vormt blijkbaar ook een risicofactor om langdurig werkloos te worden: 30.6% van de langdurige werklozen heeft geen rijbewijs, 54.9% geen auto. Bij de kortdurig werklozen is dit respectievelijk 26.0% en 45.9%. Tot slot wordt nog gekeken naar de arbeidsmarktgeschiedenis van de werklozen. Enerzijds door na te gaan hoeveel maanden men werkloos was in de twee jaar voorafgaand aan de 2007-instream, anderzijds door de beroepservaring van de individuen na te gaan. Voor deze laatste variabele beperkt de CVS-databank zich wel tot de beroepservaring in het aangegeven voorkeursberoep. Op het vlak van werkloosheidsgeschiedenis vinden we een duidelijke oververtegenwoordiging in de langdurige werkloosheid van personen die niet werkloos waren in de twee voorgaande jaren. Ook degenen die eerder reeds langdurig werkloos waren, worden relatief vaker terug langdurig werkloos. Op het vlak van beroepservaring zien we een lichte oververtegenwoordiging van degenen zonder ervaring, maar vooral de personen met meer dan vijf jaar ervaring stromen meer in in de langdurige werkloosheid. Mogelijk heeft dit te maken met het feit dat personen met meer beroepservaring doorgaans ook ouder zijn. Een regressie laat toe deze twee effecten van elkaar te onderscheiden.

## 5. De (theoretische) kans op langdurige werkloosheid

Bij statistisch profileren zal men trachten de potentieel langdurige werklozen te identificeren aan de hand van een kwantitatief verband tussen enerzijds de kans dat iemand langdurig werkloos wordt en anderzijds bepaalde kenmerken van de persoon in kwestie. Hiervoor kan men ofwel een duurmodel schatten dat de te verwachten duur van de werkloosheidsperiode aangeeft, ofwel een logistische regressie schatten die het risico op langdurige werkloosheid aangeeft. In dit rapport gaan we uit van deze tweede methode.

In tabel 5 worden de resultaten weergegeven van een waarschijnlijkheidsmodel dat de kans dat een instromer in de werkloosheid in 2007 na twaalf maanden nog steeds aanwezig zal zijn in de werkloosheidsstatistieken, relateert aan een aantal kenmerken van die persoon bij zijn of haar instroom in de werkloosheid. We schatten hierbij een apart model voor mannen en voor vrouwen.

Tabel 5 Logit kans op langdurige werkloosheid bij instroom (marginale effecten, standaard fout, z-waarde)

	Mannen			Vrouwen		
	dy/dx	std.err.	z	dy/dx	std.err.	z
<25						
25-50	0.062	0.002	33.92	0.061	0.002	30.16
≥50	0.190	0.005	38.27	0.185	0.005	37.00
Laaggeschoold						
Middengespoold	-0.004	0.002	-2.09	0.007	0.002	2.83
Hooggeschoold	-0.014	0.003	-5.36	-0.028	0.003	-10.74
Autochtoon						
Allochtoon (E.U.)	0.007	0.004	1.59	0.012	0.005	2.54
Allochtoon (niet E.U.)	0.032	0.004	8.18	0.075	0.007	10.88
Centrumstad						
Sterke functionele verstedelijking	-0.020	0.003	-6.69	-0.020	0.003	-5.93
Matige functionele verstedelijking	-0.022	0.003	-8.72	-0.018	0.003	-6.45
Zwakke functionele verstedelijking	-0.015	0.002	-6.66	-0.014	0.002	-5.60
Rijbewijs B	-0.012	0.003	-4.41	-0.004	0.003	-1.62
Auto ter beschikking	-0.022	0.002	-9.98	-0.035	0.003	-13.71
Kennis van het Nederlands	-0.035	0.004	-8.00	-0.039	0.005	-7.89
(Zeer) beperkte arbeidsgeschiktheid	0.071	0.004	16.51	0.076	0.004	17.93
BuSO	-0.007	0.005	-1.29	-0.008	0.007	-1.20
VAPH-nr	-0.003	0.006	-0.49	0.016	0.008	2.05
Niet werkloos in 2 jaren vooraf						
1-6 maanden werkloos in 2 jaren vooraf	-0.052	0.003	-20.70	-0.079	0.003	-29.22
7-12 maanden werkloos in 2 jaren vooraf	-0.033	0.003	-11.92	-0.068	0.003	-23.61
13-24 maanden werkloos in 2 jaren vooraf	-0.006	0.003	-1.94	-0.044	0.003	-14.66
Geen beroepservaring						
Tot 2 jaar beroepservaring	-0.022	0.003	-8.80	-0.026	0.003	-10.07
2 -5 jaar beroepservaring	-0.028	0.003	-10.06	-0.031	0.003	-10.48
Meer dan 5 jaar beroepservaring	-0.011	0.003	-3.95	-0.006	0.003	-2.08
Observaties	108089			102452		
Pseudo R <sup>2</sup>	0.063			0.080		

Bron: VDAB en eigen bewerkingen HIVA

De marginale effecten (dy/dx) die getoond worden in tabel 5 geven de impact van de respectievelijke variabele op de kans om langdurig werkloos te worden, terwijl gecontroleerd wordt voor alle overige opgelijste variabelen. Dit effect verschilt statistisch significant van 0 wanneer de absolute waarde van z groter is dan 2.

Een eerste zeer sterk verband in het model dat voor mannen de kans op langdurige werkloosheid schat vinden we bij de leeftijdsvariabelen. Werklozen tussen 25 en 50 hebben 6.2% meer kans om langdurig werkloos te worden dan personen jonger dan 25. Voor werklozen van 50 of ouder stijgt deze kans zelfs met 19%.

Het opleidingsniveau wordt algemeen gezien als een belangrijke voorspeller van de kans op langdurige werkloosheid. Ook uit dit model komt naar voor dat laaggeschoolden de hoogste kans hebben om langdurig werkloos te worden. Deze kans verkleint echter slechts met 0.4% voor middengespoolden. Dit zijn de personen met een diploma secundair onderwijs. Het risico dat hooggeschoolden langdurig werkloos worden, ligt 1.4% lager dan de referentiecategorie 'laagge-

schoold'. Al bij al zijn deze verschillen kleiner dan verwacht. Mogelijk heeft dit te maken met het feit dat de laaggeschoolden nu al intensiever begeleid worden door de VDAB.

Verder blijkt het risico om langdurig werkloos te worden hetzelfde voor autochtonen als allochtonen van Europese origine. Allochtonen van buiten de E.U. hebben wel 3.2% meer kans om 12 maanden na instroom nog steeds werkloos te zijn. Daarnaast blijkt ook een voldoende kennis van het Nederlands een belangrijke factor: werkzoekenden die het Nederlands machtig zijn hebben 3.5% meer kans uit te stromen uit de werkloosheid alvorens ze langdurig werkloos zijn.

Personen die in centrumsteden wonen hebben de grootste kans om langdurig werkloos te worden. Dit suggereert dat de grote tewerkstellingsopportuniteiten geen voordeel bieden aan de stedelingen.

Wat betreft mobiliteit, geeft de regressie aan dat het hebben van een rijbewijs het risico op langdurige werkloosheid met 1.2% verlaagt, het daarenboven ter beschikking hebben van een auto om te gaan werken verlaagt dit risico met 2.2%.

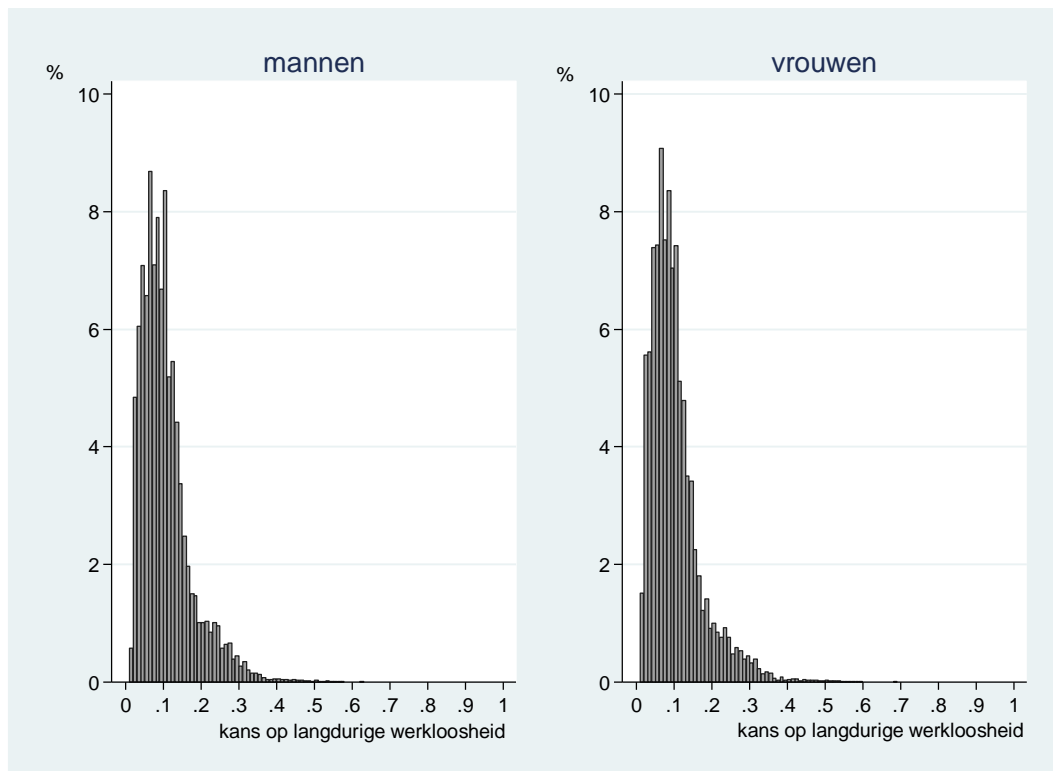
Werkzoekenden worden gedefinieerd als arbeidsgehandicapt wanneer ze behoren tot één of meer van deze drie groepen: de (zeer) beperkt arbeidsgeschikten, de ex-BuSO scholieren of de personen met een VAPH-nummer. In de regressie werden deze deelcategorieën apart opgenomen. Uit de resultaten komt naar voren dat er geen verschil is tussen ex-BuSO scholieren en de anderen, noch tussen mensen met of zonder VAPH-nummer. Personen die het label '(zeer) beperkt arbeidsgeschikt' kregen, worden daarentegen wel merkkelijk meer langdurig werkloos (+ 7.1%) dan werkzoekenden die als normaal arbeidsgeschikt beschouwd worden.

Iemands werkloosheidsgeschiedenis kan ook een belangrijke indicator zijn voor diens kans om langdurig werkloos te worden in de toekomst. In het voorspellingsmodel werd dan ook een variabele opgenomen die aangeeft of iemand in de 24 maanden voor de huidige werkloosheidsperiode niet, 1 tot 6 maanden, 7 tot 12 maanden of meer dan 13 maanden werkloos was. We zien geen verschil tussen de personen zonder werkloosheidsperiode in de afgelopen 24 maanden en degenen die langdurig werkloos waren in diezelfde periode. Mensen die in hun recent verleden een kortere tijd werkloos waren, hebben meer kans om vóór 12 maanden uit te stromen. Wanneer men eerder 1 tot 6 maanden werkloos was is dit 5.2% meer, wanneer men 7 tot 12 maanden werkloos was, 3.3%.

Tot slot wordt ook de mogelijke beroepservaring van de werkzoekenden mee in kaart gebracht. In de VDAB-data wordt enkel de ervaring in een job die in de lijn van het opgegeven voorkeursberoep ligt, opgenomen. Uit de schattingsresultaten blijkt dat het hebben van beroepservaring de kans op langdurige werkloosheid met 1 tot 2.8% verlaagt.

De resultaten voor de vrouwelijke werkzoekenden worden weergegeven in de tweede helft van tabel 5. Over het algemeen zien we dezelfde tendensen als bij de mannelijke werkzoekenden. Op het vlak van opleidingsniveau, blijkt een diploma hoger onderwijs vrouwen iets meer te beschermen tegen langdurige werkloosheid: hooggeschoolde vrouwen hebben 2.8% minder kans om langdurig werkloos te worden dan laaggeschoolde vrouwen. Daarnaast lopen allochtone vrouwen van buiten de Europese Unie 7.5% meer risico om langdurig werkloos te worden dan autochtone vrouwen. Bij de mannen was dit slechts 3.2%. Tot slot blijkt dat bij vrouwen elke eerdere werkloosheidsperiode, ook de langdurige, de kans om opnieuw langdurig werkloos te worden verlaagt.

Belangrijker is natuurlijk na te gaan wat de voorspelkracht is van het model. Met andere woorden hoe effectief is het statistische model in het aangeven van welke werkzoekenden langdurig werkloos zullen worden en welke niet. Voor iedere werkzoekende werd aan de hand van diens kenmerken een kans op langdurige werkloosheid berekend, nu moet bepaald worden boven welke grenswaarde iemand als 'hoog risico voor langdurige werkloosheid' gelabeld wordt. Grafiek 3 geeft voor mannen en vrouwen de voorspelde kans op langdurige werkloosheid.



Bron: VDAB en eigen bewerkingen HIVA  
Grafiek 3 Voorspelde kans op langdurige werkloosheid

De voorspelde kans voor langdurige werkloosheid is rechtsscheef verdeeld. Dit wil zeggen dat voornamelijk lage waarden voorkomen. Wanneer men de grenswaarde op bijvoorbeeld 50% zou stellen, zou slechts een enkeling als 'risico voor langdurige werkloosheid' gelabeld worden. De grenswaarde wordt dus best lager bepaald.

Tabel 6 geeft aan in welke mate het model erin slaagt juiste voorspellingen met betrekking tot langdurige werkloosheid te doen voor verschillende grenswaarden. In deze tabel wordt allereerst het globaal percentage correct voorspelde uitkomsten weergegeven. Daarnaast worden de sensitiviteit, dit is het aandeel correct voorspelde langdurig werklozen en de specificiteit of het aandeel correct voorspelde kortdurige (i.e. niet-langdurige) werklozen weergegeven. De tegenhanger van sensitiviteit is de verwerpingsfout met name het niet herkennen van een werkloze met een verhoogd risico, van specificiteit de doorlatingsfout of het onterecht indelen van een kortdurig werkloze bij de groep van de langdurig werklozen.



Tabel 6 Voorspelvermogen van het logit model geschat bij instroom in de werkloosheid

Grenswaarde	Globaal	Sensitiviteit	Specificiteit
<b>Mannen</b>			
Aantal (noemer)	108089	11427	96662
50%	89.4%	0.3%	100.0%
40%	89.4%	1.5%	99.8%
30%	88.9%	5.8%	98.7%
20%	85.0%	23.1%	92.3%
10%	59.1%	67.4%	58.2%
<b>Vrouwen</b>			
Aantal (noemer)	102452	11706	90746
50%	88.6%	1.0%	99.9%
40%	88.4%	4.0%	99.3%
30%	87.5%	13.3%	97.1%
20%	83.1%	30.9%	89.8%
10%	59.1%	70.9%	57.6%

Globaal= aantal juist gelabelde werklozen/ totaal aantal werklozen

Sensitiviteit= aantal langdurig werklozen die als dusdanig gelabeld worden / totaal aantal langdurig werklozen

Specificiteit= aantal kortdurig werklozen die als dusdanig gelabeld worden/ totaal aantal kortdurig werklozen

Bron: VDAB en eigen bewerkingen HIVA

Als iedereen waarvoor de voorspelde kans op langdurige werkloosheid 50% of hoger is, labelt als iemand die een hoog risico loopt om na een jaar nog steeds werkloos te zijn, is de voorspelkracht van dit model zeer pover. Van de mannen die effectief langdurig werkloos werden wordt dan slechts 0.3% als dusdanig herkend, bij de vrouwen geldt dit voor 1%. In absolute aantallen wil dit dus zeggen dat slecht 39 van de uiteindelijk 11 427 langdurig werkloze mannen en 115 van de 11 706 langdurig werkloze vrouwen juist gelabeld worden. Een grenswaarde van 50% is, zoals ook al bleek uit grafiek 3, wellicht aan de hoge kant. Wanneer een kleinere grenswaarde vastgelegd wordt, zien we de sensitiviteit stijgen. Meer en meer effectief langdurig werklozen worden als dusdanig herkend door het model. Voor de mannen is dit bij een grenswaarde van 40% 1.5%, bij een grenswaarde van 30% 5.8%, bij een grenswaarde van 20% 23.1% en bij een grenswaarde van 10%, 67.4%. Voor de vrouwen stijgt de sensitiviteit voor deze grenswaarden van 4.0% tot 70.9%. Deze sensitiviteit stijgt echter tegen een hoge kost. Door meer en meer mensen te labelen als 'potentieel langdurig werkloos' zal men er wel in slagen de eigenlijke langdurige werklozen beter op te sporen, maar ook personen die nooit langdurig werkloos worden, worden aldus gelabeld. Dit wordt duidelijk wanneer we de specificiteit bekijken. Deze waarde geeft immers weer in welke mate de kortdurig werklozen ook aldus gelabeld worden. De grote daling in specificiteit ligt tussen de grenswaarde van 20% en 10% waar de specificiteit van de mannen daalt van 92.3% naar 58.2% en die bij de vrouwen van 89.8% naar 57.6%.

De grenswaarde is een waarde die door het beleid moet bepaald worden op basis van doelstellingen en budgetten. Zoals tabel 6 aangeeft zal het aandeel juist gelabelde langdurig werklozen stijgen, naarmate men de grenswaarde om iemand met een groot risico op langdurige werkloosheid te identificeren verlaagt. Tegelijkertijd zullen zo echter ook steeds meer personen die geen reëel risico lopen om langdurig werkloos te worden wel als dusdanig gelabeld worden. Wanneer we ervan uitgaan dat men de personen met een grote kans om langdurig werkloos te worden gaat (eer-

der) gaat benaderen met activerende maatregelen en acties, zal er sprake zijn van overbodige kosten, de zogenaamde deadweight kosten.

De voorspelwaarde van het model kan verbeteren door het gebruik van betere data. Men kan proberen de momenteel niet-geobserveerde heterogeniteit te vatten door het stellen van bijkomende vragen bij de inschrijving als werkzoekende. Zo gaven andere studies in binnen- en buitenland (Desmet 2010, O'Connell et al. 2010) aan dat de gezinssituatie, het inkomen van de partner, de motivatie en de algemene gezondheid een invloed kunnen hebben op het al dan niet langdurig werkloos worden. Hoewel in de praktijk consultants vaak naar deze bijkomende en vaak zachtere gegevens peilen, zijn deze momenteel niet beschikbaar voor analyse.

Een andere manier om de voorspelwaarde van het model te verbeteren is door de kans op langdurige werkloosheid niet bij de instroom in de werkloosheid trachten te voorspellen, maar pas wanneer men enige tijd werkloos is (Bollens et al. 2003). Tabel 7 geeft voor de mannelijke werkzoekenden de kans op langdurige werkloosheid na drie en zes maanden werkloosheid.

Tabel 7 Logit kans op langdurige werkloosheid bij mannen na 3 en 6 maanden werkloosheid (marginale effecten, standaard fout, z-waarde)

Mannen	Na 3 maanden			Na 6 maanden		
	dy/dx	std.err	z	dy/dx	std.err	z
<25						
25-50	0.117	0.005	23.60	0.131	0.009	14.25
≥50	0.309	0.009	34.87	0.336	0.012	27.22
Laaggeschoold						
Middengeschoold	-0.002	0.005	-0.35	-0.002	0.007	-0.23
Hooggeschoold	-0.028	0.006	-4.42	-0.023	0.010	-2.40
Autochtoon						
Allochtoon (E.U.)	-0.003	0.010	-0.29	-0.013	0.014	-0.94
Allochtoon (niet E.U.)	0.047	0.008	5.88	0.050	0.011	4.47
Centrumsteden						
Sterke functionele verstedelijking	-0.027	0.007	-3.82	-0.024	0.011	-2.24
Matige functionele verstedelijking	-0.034	0.006	-5.69	-0.036	0.009	-3.99
Zwakke functionele verstedelijking	-0.021	0.005	-4.10	-0.016	0.008	-2.04
Rijbewijs B	-0.020	0.006	-3.42	-0.024	0.009	-2.63
Auto ter beschikking	-0.027	0.005	-5.39	-0.026	0.008	-3.42
Kennis van het Nederlands	-0.028	0.010	-2.83	-0.002	0.014	-0.13
(Zeer) beperkte arbeidsgeschiktheid	0.122	0.010	12.10	0.136	0.016	8.67
BuSO	0.005	0.012	0.40	0.016	0.018	0.92
VAPH-nr	-0.015	0.015	-1.02	0.003	0.023	0.13
0 mnd werkloos in 2 jaren vooraf						
1-6 maanden werkloos in 2 jaren vooraf	-0.071	0.006	-12.19	-0.054	0.009	-6.04
7-12 maanden werkloos in 2 jaren vooraf	-0.040	0.006	-6.67	-0.019	0.009	-2.04
13-24 maanden werkloos in 2 jaren vooraf	0.003	0.006	0.46	0.020	0.009	2.40
Geen beroepservaring						
Tot 2 jaar beroepservaring	-0.036	0.006	-6.35	-0.035	0.008	-4.14
2-5 jaar beroepservaring	-0.051	0.006	-7.94	-0.054	0.010	-5.47
Meer dan 5 jaar beroepservaring	-0.012	0.006	-1.98	-0.004	0.009	-0.48
Observaties	43066			24411		
Pseudo R <sup>2</sup>	0.047			0.036		

Bron: VDAB en eigen bewerkingen HIVA

Uit tabel 7 blijken dezelfde statistische verbanden die hoger reeds besproken werden. Door het identificatiemoment uit te stellen hopen we echter dat de voorspellingsprestaties van het model verbeteren. Tabel 8 geeft een overzicht van het voorspelvermogen op de verschillende momenten.

Tabel 8 Voorspelwaarde van het model bij instroom, na 3 en 6 maanden

	Globaal	Sensitiviteit	Specificiteit
Bij instroom			
Aantal (noemer)	108089	11427	96662
Grenswaarde = 20%	85.0%	23.1%	92.3%
Na 3 maanden			
Aantal (noemer)	43066	11427	31639
Grenswaarde = 30%	67.7%	41.6%	77.1%
Na 6 maanden			
Aantal (noemer)	24411	11427	12984
Grenswaarde = 45%	58.1%	59.4%	56.9%

Globaal= aantal juist gelabelde werklozen/ totaal aantal werklozen

Sensitiviteit= aantal langdurig werklozen die als dusdanig gelabeld worden / totaal aantal langdurig werklozen

Specificiteit= aantal kortdurig werklozen die als dusdanig gelabeld worden/ totaal aantal kortdurig werklozen

Bron: VDAB en eigen bewerkingen HIVA

De grenswaarden om een individu te beschouwen als iemand met een hoog risico om langdurig werkloos te worden, wordt bepaald aan de hand van het feitelijke aandeel langdurig werklozen binnen de steekproef. Bij de instroom wordt de grenswaarde zo vastgelegd op 20%, na drie maanden op 30% en na zes maanden op 45%. We zien dat de sensitiviteit gevoelig toeneemt. Waar we bij de instroom amper een kwart van de langdurig werklozen (2 636) konden identificeren, is dit na drie maanden voor vier op de tien (4 754) en na zes maanden voor zes op de tien (6 793) het geval. De specificiteit op haar beurt daalt, wat impliceert dat de doorlatingsfout toeneemt. Hierbij moet echter opgemerkt worden dat het hier om een relatieve maat gaat en dat de deadweight in absolute termen niet stijgt. Bij instroom worden immers 7 467 of 7.7% kortdurig werklozen foutief als potentieel langdurig werkloos gelabeld, na drie maanden is dit percentage weliswaar opgelopen tot 22.9% maar zijn dit slechts 7 243 personen en na zes maand gaat het nog om 5 599 werklozen.

## 6. Conclusie

Om de schaarse middelen voor de activering van werklozen efficiënt te kunnen inzetten, zou het interessant zijn om kort na de instroom in de werkloosheid reeds een idee te hebben van ieders afstand tot de arbeidsmarkt. Statistisch diagnosticeren wordt in deze context vaak gezien als een manier om eenvoudig en snel een antwoord te krijgen op de vraag: 'Wie heeft de hoogste kans om langdurig werkloos te blijven?'. Concreet gaat men op zoek naar een kwantitatief verband tussen enerzijds de kans dat iemand langdurig werkloos wordt en anderzijds de kenmerken van de persoon in kwestie. In vele landen gebruikt men al zulk regressiemodel of beslissingsboom om grote groepen instromers in te delen in groepen naargelang hun risico op langdurige werkloosheid. De bedoeling is dan uiteraard om de dienstverlening maximaal aan te passen aan de noden van de werkzoekende. Personen met een grote kans om langdurig werkloos te worden, zouden dan sneller intensieve begeleiding of opleiding krijgen, terwijl overbodige acties en kosten voor degenen die snel uitstromen vermeden kunnen worden.

In deze paper wordt een theoretische kans op langdurige werkloosheid voor mannen en vrouwen afzonderlijk berekend op basis van de CVS-databank van de VDAB. Deze databank wordt aangevuld bij de registratie als niet werkende werkzoekende en bevat voornamelijk informatie over de zogenaamde harde kenmerken van de werkzoekende. We vonden in beide modellen statistisch significante effecten van leeftijd, opleidingsniveau, etniciteit, arbeidshandicap, woonplaats, mobiliteit en arbeidsmarktgeschiedenis.

Belangrijker is echter om te kijken naar de voorspelkracht van het model. Hiervoor werden de voorspelde waarden vergeleken met de eigenlijke uitkomst na een jaar. Cruciaal hierbij is de keuze van de grenswaarde, vanaf wanneer beschouwen we iemand als 'een werkloze met een verhoogde kans op langdurige werkloosheid'. Is dit wanneer de voorspelde kans op langdurige werkloosheid meer dan 50% is, of moet deze grens hoger of lager gesteld worden? Wanneer we ons baseren op het werkelijke aandeel dat langdurig werkloos wordt<sup>5</sup>, moet de grenswaarde ongeveer op 20% gesteld worden. Ieder met een theoretische kans op langdurige werkloosheid van meer dan 20%, wordt dan als 'hoog risico op langdurige werkloosheid' bekeken. Op basis van het regressiemodel worden bij deze grenswaarde 22 973 (10 103 mannen en 12 870 vrouwen) van de 210 541 instromers van 2007 als risicovol geclassificeerd. Voor deze personen zou men dus bij voorkeur snel een traject moeten opstarten. De sensitiviteit, dit zijn de langdurig werklozen die als dusdanig correct geclassificeerd werden, is 23.1% bij de mannen en 30.9% bij de vrouwen. Concreet gaat het dan om 2 636 van de 11 427 mannelijk langdurige werklozen en om 3 612 van de 11 706 vrouwelijk langdurige werklozen voor wie terecht zulk traject opgestart wordt. De overige 16 885 (8 791 mannen en 8 094) langdurig werklozen worden hierbij initieel over het hoofd gezien. De specificiteit, de feitelijke niet langdurig werklozen die correct ingedeeld werden, ligt bij de grenswaarde van 20% op respectievelijk 92.3% en 89.8%. Het complement van deze waarden zijn de personen die vóór twaalf maanden werkloosheid uitstroonden, maar toch als potentieel langdurig werkloos geclassificeerd werden. Het gaat hier om 7 467 mannen en 9 258 vrouwen. Voor hen zal men dus allerlei onnodige activeringsacties opstarten, naar dit fenomeen verwijst men doorgaans met de term deadweight kosten. Het is duidelijk dat deze waarden steeds bepaald worden door de keuze van de grenswaarde, waarbij men moet beseffen dat men een hogere sensitiviteit steeds betaald met een lagere specificiteit en omgekeerd.

Om de sensitiviteit en specificiteit tegelijkertijd te verhogen, moet de voorspelkracht van het model verbeteren. Dit kan door het gebruik van betere gegevens. Men zou de dataset kunnen verrijken met bijkomende relevante kenmerken. Zo zien we in het buitenland meer en meer pogingen om de zachte kenmerken van werkzoekenden te meten en te verwerken in de statistische analyse. In Vlaanderen werden eerder reeds vragenlijsten ontwikkeld om de CVS-gegevens aan te vullen<sup>6</sup>, deze gegevens zijn evenwel niet beschikbaar voor analyse in een dataset. Een andere manier om meer exact te kunnen voorspellen wie een hoog risico loopt om langdurig werkloos te worden, is de identificatie van de mogelijk langdurig werklozen niet dadelijk bij de instroom in de werkloosheid te organiseren, maar een aantal maanden uit te stellen. In het rapport voorspelden we de kans op langdurige werkloosheid voor de mannelijke werklozen na drie en zes maanden werkloosheid. Wanneer we de grenswaarden opnieuw vastleggen op het feitelijk gemiddelde (i.e. 30% bij drie maanden en 45% bij zes maanden) vinden we een sensitiviteit van respectievelijk 41.6% en 59.4% en een specificiteit van 77.1% en 56.9%. De mate waarin langdurig werklozen als dusdanig geclassificeerd worden gaat er met ander woorden behoorlijk op vooruit. De specificiteit daalt wel, maar men

<sup>5</sup> Bij de mannelijke instromers van 2007 is dit 10.6%, bij de vrouwen 11.4%.

<sup>6</sup> In het kader van INDICATE, een EQUAL-project, werd een lijst van een twintigtal vragen ontwikkeld die moeten helpen de kans op langdurige werkloosheid beter in te schatten. Deze lijst zit momenteel in de toolbox van de VDAB-consulent.

mag niet uit het oog verliezen dat deze groep in absolute termen wel steeds kleiner wordt. Wanneer bij instroom en een grenswaarde van 20% nog 7 467 feitelijk kortdurig werklozen nog foutief gelabeld worden, zijn dit er na drie maanden bij een grenswaarde van 30% nog 7 243 en na zes maanden bij een grenswaarde van 45% nog maar 5 599.

Algemeen wordt statistische profilering beschreven als een instrument dat op een snelle en eenvoudig interpreteerbare manier een grote groep van werklozen kan indelen in degenen met een hoog risico op langdurig werkloosheid en degenen met een lager risico. Bovendien heeft deze manier van diagnosticeren als grote voordeel dat de beschikbare informatie voor alle werkzoekenden op dezelfde manier gebruikt en geëvalueerd wordt. Het gaat dus om een erg objectief oordeel van iemands afstand tot de arbeidsmarkt. Daarnaast moeten toch ook enkele kanttekeningen gemaakt worden bij dit diagnose-instrument. Ten eerste blijft de kans om personen fout te classificeren steeds vrij hoog. Zoals eerder uitvoerig toegelicht zal men hier de afweging moeten maken tussen enerzijds het te weinig identificeren van de potentieel langdurig werklozen, waardoor het instrument eigenlijk zijn doel voorbij schiet en anderzijds het te veel identificeren van potentieel langdurig werklozen, waardoor men te maken krijgt met hoge deadweight kosten. Verder mag ook de kost van een degelijk statistisch model niet onderschat worden. Geavanceerde statistische diagnose-instrumenten vereisen immers grote initiële kosten. We kunnen hier denken aan de kosten voor het identificeren en verzamelen van relevante data, de kosten voor het ontwikkelen van de software en de kosten voor het opleiden van de consultants om het systeem juist te gebruiken en de resultaten te interpreteren. Tot slot wordt ook soms verwezen naar het locking-in effect. Door een zeer vroegtijdige opvolging, identificatie en activering van bepaalde werkzoekenden zijn deze misschien minder geneigd om snel, autonoom op zoek te gaan naar werk.

Ten slotte moet er nog aan herinnerd worden dat het identificeren van potentieel langdurig werkzoekenden natuurlijk niet op zich staat, maar slechts nuttig is, wanneer men deze personen nadien ook kan activeren aan de hand van effectieve acties.

## Bibliografie

- Black D., Smith J., Berger M. & Noel B. (2003), Is the threat of reemployment services more effective than the services themselves? Evidence from random assignment in the UI system, *The American Economic Review*, Vol 93, No 4, p.1313-1327.
- Bollens J., Heylen V. & Vos S. (2003), Screening van de arbeidsmarktrijpheid tegen de achtergrond van een preventief activerend arbeidsmarktbeleid, HIVA-K.U.Leuven, Leuven.
- Brimrose J., Barnes S.A., Brown A., Hasluck C. & Behle H. (2007), Skills diagnostics and screening tools: a literature review. Department for Work and Pensions Research report No 459, Warwick Institute for Employment Research, Warwick.
- Collewet M., Gravesteyn J., de Koning J. (2009), Diagnose-instrumenten bij re-integratie: ervaringen uit zeven landen over de afgelopen 20 jaar, Paper voor de Nederlandse Arbeidsmarktdag, 14 oktober 2009, SER, Den Haag.
- Debauche E. & Georges N. (2007), Construire un modèle de profilage des demandeurs d'emploi: défi statistique ou défi politique?, Centre d'études de l'emploi, Noisy-le-Grand.
- Desmet L. (2010), Een onderzoek naar de determinanten van de uitstroom naar werk, Directie Statistiek en Studies RVA.
- Hasluck C. (2008), The use of statistical profiling for targeting employment services: The international experience, in Di Domenico G. & Spattini S. (eds.), *New European approaches for long-term unemployment: What role for public employment services and what market for private stakeholders?* Kluwer Law International, Nederland.
- Heylen V. (2010), Duration dependence in Flemish unemployment data, WSE Report, Leuven.
- Konle-Seidl R. (2011), Profiling systems for effective labour market integration. Use of profiling for resource allocation, action planning and matching. Mutual Learning Programme for Public Employment Services.
- Mosley H. (2010), Reform of placement services. Peer Review on "Systematic Preventive Integration Approach (Support) for Jobseekers and Unemployed", Mutual Learning Programme, Germany, 28-29 October.
- Nationale Bank van België (2011), Jaarverslag 2010, Brussel.
- O'Connell P., McGuinness S. & Kelly E. (2010), A statistical profiling model of long-term unemployment risk in Ireland, Working Paper no.345, ESRI.
- O'Leary C. & Eberts R. (2010), Employment and training policy in the United States during the economic crisis. Upjohn Institute Working Paper, No 10-161.
- Payne C. & Payne J. (2000), Early identification of the long-term unemployed, PSI Research Discussion Paper, London.
- Riipinen T. (2011), Risk profiling of long-term unemployed in Finland, Presentation for the PES Dialogue Conference, 8<sup>th</sup> of September 2011, Brussels.