



Wetenschappelijk Onderzoek- en
Documentatiecentrum
Ministerie van Justitie en Veiligheid

Cahier 2018-4

Recidive na Tools4U

Vergelijkend recidiveonderzoek naar het effect van Tools4U

T. van der Stouwe (UvA)
S. Verweij (WODC)
J.J. Asscher (UvA, UU)
G. Weijters (WODC)
G.J.J.M. Stams (UvA)

Cahier

De reeks Cahier omvat de rapporten van onderzoek dat door en in opdracht van het WODC is verricht.

Opname in de reeks betekent niet dat de inhoud van de rapporten het standpunt van de Minister van Justitie en Veiligheid weergeeft.

Inhoud

Afkortingen — 4

Samenvatting — 5

1 Inleiding — 9

- 1.1 Tools4U — 9
- 1.2 Context — 10
- 1.3 Onderzoek naar Tools4U tot nu toe — 11
- 1.4 Recidive onderzoek — 13

2 Methode — 15

- 2.1 Gebruikte gegevensbronnen — 15
- 2.2 Samenstelling onderzoeksgroepen — 15
- 2.3 Recidivemetingen — 18
- 2.4 Statistische analyses — 19

3 Beschrijving onderzoeksgroepen — 21

- 3.1 Achtergrondkenmerken Tools4U-deelnemers — 21
- 3.2 Vergelijkbaarheid Tools4U-deelnemers en Controlegroep 1 — 23
- 3.3 Vergelijkbaarheid Tools4U-deelnemers en Controlegroep 2 — 23

4 Resultaten — 25

- 4.1 Recidiveprevalentie — 25
- 4.2 Ernst eerste recidivezaak — 27
- 4.3 Kenmerken die van invloed zijn op effectiviteit — 28

5 Conclusie en discussie — 31

- 5.1 Conclusie — 31
- 5.2 Discussie — 33

Summary — 35

Literatuur — 39

Bijlagen

- 1 Samenstelling Begeleidingscommissie — 43
- 2 Beschrijving analysetechnieken — 45

Afkortingen

dde	detentiedag-equivalenten
JDS	Justitieel Documentatiesysteem
JustId	Justitiële Informatiedienst
LVB	licht verstandelijke beperking
OBJD	Onderzoeks- en Beleidsdatabase Justitiële Documentatie
OM	Openbaar Ministerie
MSI	Misdaad Straf Index
PSM	Propensity Score Matching
RvdK	Raad voor de Kinderbescherming
UvA	Universiteit van Amsterdam
WODC	Wetenschappelijk Onderzoek- en Documentatiecentrum

Samenvatting

Tools4U is een leerstraf gericht op het vergroten van sociale en cognitieve vaardigheden voor jongeren tussen de 12 en 18 jaar met een middelgroot recidiverisico. Eerder onderzoek naar de uitvoering en effectiviteit van deze interventie wat betreft cognitieve en sociale vaardigheden lieten overwegend positieve effecten zien. Tools4U voldoet bovendien grotendeels aan de *What Works*-beginselen waaraan een interventie moet voldoen om effectief te zijn in het verminderen van recidive.

In de eerdere studie naar de effecten van Tools4U op recidive werden 115 Tools4U jongeren vergeleken met 108 jongeren die een jaar later een andere taakstraf kregen. Deze jongeren waren op basis van *propensity score matching* (PSM) aan de Tools4U-jongeren gematcht. Er werd geen significant verschil in recidive gevonden tussen de Tools4U-groep en de controlegroep na zes maanden, twaalf maanden, de langst mogelijke observatietermijn die voor alle jongeren beschikbaar was (1,46 jaar) en de maximale observatie duur (2,76 jaar). Verder werden er geen modere-rende effecten gevonden voor leeftijd, geslacht, herkomst en delict geschiedenis. Er werd wel een marginaal significant verschil gevonden in geweldsdelicten: als Tools4U-jongeren recidiveerden, dan deden zij dit minder vaak met een geweldsdelict dan jongeren in de controlegroep.

Omdat de eerdere studie recidive na een relatief korte termijn (1,5 jaar) onderzocht, wordt in de huidige studie een recidivemeting uitgevoerd met een langere recidivetermijn. De eerdere studie maakte gebruik van een controlegroep die ruim een jaar later dan de Tools4U-groep geworven was, waardoor er een verschil was in de observatietermijn tussen de beide groepen. In de huidige studie is daarom daarnaast een extra controlegroep samengesteld die in dezelfde periode als de Tools4U-groep een taakstraf ontving. Bovendien zijn er aanvullende analyses gedaan naar de invloed van sociale- en cognitieve vaardigheden bij de voormeting.

De volgende onderzoeksvragen worden in het huidige onderzoek beantwoord:

- In welke mate is Tools4U effectief in het voorkomen van recidive? Op basis van:
 - prevalentie;
 - ernst eerste recidivedelict.
- Voor welke jongeren is Tools4U meer of minder effectief? Op basis van:
 - leeftijd, geslacht, herkomst, delict geschiedenis;
 - vaardigheidstekorten voorafgaand aan de training.

Methode

In dit onderzoek zijn gegevens uit verschillende bronnen met elkaar gecombineerd. De primaire gegevensbron is een bestand samengesteld in het kader van het onderzoek naar de effecten van Tools4U op cognitieve en sociale vaardigheden. Aan het bestand zijn gegevens uit de Onderzoeks- en Beleidsdatabase Justitiële Documentatie (OBJD) gekoppeld. Met behulp van de gegevens in de OBJD kan per jeugdige de strafrechtelijke carrière in kaart worden gebracht.

Het onderzoek heeft betrekking op 115 jeugdige daders die in 2012 een Tools4U-training hebben doorlopen. De uitkomsten van de Tools4U-deelnemers worden vergeleken met die van twee controlegroepen. Controlegroep 1 is met PSM samengesteld in het eerdere effectonderzoek van Tools4U en bevat 108 jeugdige justitiabelen die in 2013 of 2014 een andere leer- of werkstraf kregen opgelegd. Controlegroep 2 is samengesteld met behulp van PSM op de volgende tien achtergrond-

kenmerken uit de OBJD: sekse, leeftijd bij uitstroom, geboorteland, delict categorie, arrondissement, modaliteit, ernst uitgangszaak, leeftijd eerste strafzaak, aantal eerdere strafzaken en strafzaakdichtheid. Uiteindelijk bestaat Controlegroep 2 uit 115 personen.

De recidivemetingen zijn uitgevoerd volgens de gestandaardiseerde werkwijze van de WODC-Recidivemonitor. In dit onderzoek wordt zowel naar de recidiveprevalentie als naar de ernst van het eerste recidivedelict gekeken.

Resultaten

Er zijn weinig verschillen gevonden tussen de Tools4U-deelnemers en de controlegroepen wat betreft achtergrondkenmerken. Alleen Controlegroep 1 wijkt licht af van de Tools4U-deelnemers wat betreft het type delict, de leeftijd bij de eerste strafzaak en het arrondissement waarin de straf werd opgelegd.

Het recidiveonderzoek wijst uit dat een jaar na de afronding van Tools4U 17% van de deelnemers heeft gerecidiveerd. Na twee en drie jaar loopt dit percentage op naar respectievelijk 27% en 41%. De één- en tweejarige recidivepercentages van de Controlegroep 1 liggen hoger. Na één jaar recidiveert 25% van de jeugdigen in Controlegroep 1 en de tweejarige recidiveprevalentie bedraagt 33%. Het driejarige recidivepercentage van Controlegroep 1 ligt met 38% net iets lager dan de recidiveprevalentie van de Tools4U-deelnemers. Van de jeugdigen in Controlegroep 2 recidiveert 20% in het eerste jaar, 32% in het tweede jaar en 38% in het derde jaar. De verschillen tussen recidivecurves van de Tools4U-deelnemers en de beide controlegroepen zijn niet statistisch significant.

De gemiddelde ernst van de eerste recidivezaak van de Tools4U-deelnemers blijkt net iets hoger te liggen dan de gemiddelde ernst van de eerste recidivezaak van de beide controlegroepen. Ook deze verschillen zijn echter niet statistisch significant. In het huidige recidiveonderzoek hebben we dus geen bewijs kunnen vinden van de effectiviteit van Tools4U.

Als laatste hebben we gekeken of de deelname aan Tools4U voor jeugdigen met bepaalde kenmerken een verschillend effect heeft gehad. Bij deze analyse wordt alleen Controlegroep 1 betrokken omdat voor Controlegroep 2 geen informatie beschikbaar is over vaardigheidstekorten en we vooral geïnteresseerd zijn of er indicaties zijn dat Tools4U effectief is voor jeugdigen bij wie vooraf bepaalde vaardigheidstekorten gemeten zijn. Uit de moderatoranalyse blijkt dat jeugdigen in de controlegroep met denkfouten op het gebied van goedpraten of verkeerd benoemen van antisociaal gedrag meer kans hebben op recidive dan Tools4U-deelnemers met problemen op dit gebied. Dit lijkt een aanwijzing te zijn dat Tools4U het recidiveverhogende effect van deze denkfouten geneutraliseerd heeft. Geen van de andere vaardigheidstekorten of achtergrondkenmerken liet een verschillend effect zien voor Tools4U-deelnemers ten opzichte van de jongeren in de controlegroep.

Conclusie

Op basis van dit onderzoek kan niet geconcludeerd worden dat Tools4U meer of minder effectief is dan een andere leer- of werkstraf in het voorkomen van recidive. Deze uitkomsten zijn onverwacht, omdat Tools4U op basis van eerder onderzoek voldoet aan bijna alle beginselen waar aan voldaan moet worden om effectief te kunnen zijn. De uitkomsten zijn bovendien niet in overeenstemming met eerdere overzichtsstudies naar sociale-vaardigheidstrainingen en justitiële jeugdinterventies, die kleine langetermijneffecten vonden. In deze studies was programma-integriteit

echter vaak niet gemeten, waardoor de inzet van andere technieken mogelijk tot de effecten heeft geleid. Bovendien vergeleken de meeste onderzoeken het effect van de sociale-vaardigheidstraining met een controlegroep die geen behandeling kreeg, wat een gevolg kan zijn van het feit dat het beter is om iets te doen dan niets te doen (i.e., effect van aandacht). Tot slot richt Tools4U zich op een relatief 'lichte' doelgroep en is daarom weinig intensief, met weinig ruimte voor multimodaal aangrijpen. Daarom mogen wellicht dus ook maar kleine effecten verwacht worden, die binnen een individuele studie zoals de huidige moeilijk te detecteren zijn. Bovendien zijn er specifiek in Nederland nog maar weinig interventies waarvan is aangetoond dat zij effectief zijn in het verminderen van recidive in vergelijking met behandeling zoals gebruikelijk.

Wel werd gevonden dat jongeren in de controlegroep met meer denkfouten met betrekking tot 'goedpraten/verkeerd benoemen', vaker recidiveerden dan Tools4U-deelnemers. Deze bevinding komt deels overeen met de uitkomsten van het onderzoek naar de effecten van Tools4U op de sociale en cognitieve vaardigheden: meer Tools4U-deelnemers waren vooruitgegaan op het probleemgebied 'goedpraten/verkeerd benoemen' dan jeugdigen uit de controlegroep, alleen was dit verschil slechts marginaal significant.

De huidige uitkomsten moeten met enige voorzichtigheid geïnterpreteerd worden, omdat de deelnemers niet willekeurig aan de behandelconditie toegewezen waren en er dus andere onbekende factoren van invloed kunnen zijn geweest op de verschillen tussen de behandelgroepen. Daarnaast is het huidige aantal deelnemers niet groot genoeg om kleine significante verschillen te kunnen detecteren.

De uitkomsten van dit rapport betekenen voor de praktijk dat de effectiviteit van Tools4U minder groot of minder aannemelijk is dan gedacht werd. De mate waarin Tools4U op basis van de eerdere onderzoeken voldoet aan de *What Works*-beginselen biedt echter wel aanknopingspunten om de indicatie voor of uitvoering van Tools4U op bepaalde aspecten te verbeteren. Verder zouden in een volgende studie ook bredere uitkomstmaten onderzocht moeten worden, zoals (zelfgerapporteerde) politiecontacten, dagbesteding, opleiding en werk om vast te stellen of Tools4U wel tot vooruitgang leidt op die gebieden, bij voorkeur bij een groter aantal deelnemers, zodat ook kleine(re) effecten of effecten op andere relevante uitkomstgebieden gevonden kunnen worden.

1 Inleiding

Sinds de officiële invoering van taakstraffen als hoofdstraf maakt de sociale-vaardigheidstraining onderdeel uit van het leerstraffenpakket voor delinquentie jongeren in Nederland. Deze training past goed in het veranderende klimaat van vooral preventief in plaats van punitief ingrijpen bij kwetsbare jongeren (Tweede Kamer der Staten-Generaal, 1996). De oorspronkelijke training sociale vaardigheden voor jongeren met gedragsproblemen van het Amsterdamse Paedologisch Instituut, die in de jaren tachtig van de vorige eeuw als een van de eerste leerprojecten als alternatieve sanctie werd toegepast, is sindsdien doorontwikkeld tot de huidige versie: Tools4U (Albrecht & Spanjaard, 2011). In het kader van evidence based werken is (onderzoek naar) de effectiviteit van leerstraffen en dus van deze sociale-vaardigheidstraining de afgelopen decennia steeds belangrijker geworden. Daarom is Tools4U de afgelopen jaren veelvuldig onderzocht: er vonden een procesevaluatie (Bosma, Asscher, Van der Laan & Stams, 2011), doeltreffendheidsonderzoek (Van der Stouwe, Asscher & Stams, 2013) en een effectstudie naar de effecten op cognitieve en sociale vaardigheden (Van der Stouwe, Asscher, Stams, Hoeve & Van der Laan, 2015) en recidive (Van der Stouwe, Asscher, Hoeve, Van der Laan & Stams, 2017) plaats, waarin Tools4U overwegend positieve effecten liet zien. Omdat de eerdere recidivestudie op relatief korte termijn (1,5 jaar) recidive onderzocht, wordt in de huidige studie opnieuw een recidivemeting uitgevoerd met een langere recidive-termijn (3 jaar). Verder werd in de eerdere studie gebruikgemaakt van een controlegroep die ruim een jaar later dan de Tools4U-groep geworven was, waardoor een verschil bestaat in de observatietermijn van de beide groepen. In de huidige studie is daarom daarnaast een extra controlegroep samengesteld die in dezelfde periode als de Tools4U-groep een taakstraf ontving. Bovendien zijn er aanvullende analyses gedaan naar de invloed van sociale en cognitieve vaardigheden bij de voormeting.

1.1 Tools4U

Tools4U is een sociale-vaardigheidstraining die als leerstraf opgelegd kan worden door de kinderrechter of officier van justitie (OM-afdoening). Deze ambulante training bestaat uit acht tot twaalf individuele wekelijkse bijeenkomsten van 1,5 uur die – in het geval van problemen met toezicht en probleemoplossing van de ouders – aangevuld kunnen worden met twee bijeenkomsten voor de ouders en twee bijeenkomsten voor ouders en de jongere samen (Albrecht & Spanjaard, 2011). De training richt zich op jongeren tussen de 12 en 18 jaar met een middelgroot recidiverisico¹, waarbij cognitieve- en/of sociale-vaardigheidstekorten een rol speelden bij het plegen van het delict, die bereid zijn om aan de training deel te nemen en waarbij individuele training de voorkeur heeft boven een groepstraining. Er mogen niet te veel of te weinig vaardigheidstekorten in relatie tot het delict zijn, er mag geen sprake zijn van problematisch functioneren of veel verschillende gebieden of ernstige symptomen van psychopathologie die het delict veroorzaakt hebben en er mag geen sprake zijn van een licht verstandelijke beperking (LVB). Bovendien mag

¹ Zoals bepaald met het risico-inschattingsinstrument Prescreen, op basis van demografische gegevens, ernst van het delict en delict geschiedenis.

de jongere niet al op jonge leeftijd begonnen zijn met antisociaal gedrag en mag het risico op geweld tegen personen niet te hoog zijn.

Tijdens de training wordt er gewerkt aan het verbeteren van cognitieve en sociale vaardigheden die in relatie stonden tot het plegen van het delict (Albrecht & Spanjaard, 2011; Spanjaard, 2013). Voorbeelden van vaardigheden die getraind worden, zijn prosociale contacten aangaan, omgaan met groepsdruk, sociaal perspectief nemen, probleemoplossingsvaardigheden, zelfbeheersing en omgaan met emoties en kritisch redeneren (Spanjaard, Regterschot-Von Lindheim, Groenhuijsen & Bolt, 2012). Aangenomen wordt dat het verbeteren van deze vaardigheden de jongere helpt om in een volgende risicosituatie een sociale(re) keuze te maken en dat daarmee recidive voorkomen wordt.

1.2 Context

Hoewel er een veelvoud aan studies is uitgevoerd naar de effecten van sociale-vaardigheidstrainingen, is er tot nu toe nog geen overzichtsstudie beschikbaar die specifiek het effect van sociale-vaardigheidstrainingen op recidive heeft onderzocht. Wel zijn sociale-vaardigheidstrainingen als interventiegroep geïncorporeerd binnen de categorie vaardigheidstrainingen in een overzichtsstudie naar de effecten van justitiële jeugdinterventies op officiële recidive. In deze studie lieten sociale-vaardigheidstrainingen na een jaar een algemene recidive reductie van 13% zien; een klein positief effect (Lipsey, Howell, Kelly, Chapman & Carver, 2010; Lipsey, 2009).

Er zijn verder ook overzichtsstudies uitgevoerd die zich wel specifiek richten op sociale-vaardigheidstrainingen, maar daarbij ingaan op een verscheidenheid aan uitkomstmaten (Ang & Hughes, 2002; Cook et al., 2008; Maag, 2006). In deze overzichtsstudies werden middelgrote effecten meteen na de training en kleine effecten zo'n zes maanden na de interventie gevonden. In deze studies werd echter een bredere doelgroep van jongeren met gedragsproblemen geïncorporeerd (niet alleen delinquente jongeren) en werd de sociale-vaardigheidstraining in veel onderzoeken vergeleken met een wachtlijst of placebo controlegroep. Bovendien werden voor de *overall*-effecten een diversiteit aan uitkomsten samengevoegd.² Hoewel recidive daarbij geïncorporeerd werd, waren de meeste uitkomsten 'zachtere' maten van gedrag en vaardigheden.

Interventies zijn aantoonbaar effectiever in het verminderen van recidive indien aan een zestal *What Works*-beginselen is voldaan (Andrews & Bonta, 2010a; Bonta & Andrews, 2007; Van der Laan, 2004). Veel van de interventies in bovengenoemde overzichtsstudies voldeden niet aan al deze kenmerken. Het gaat daarbij ten eerste om het *risicobeginsel*, dat stelt dat de intensiteit van de interventie aangepast moet worden aan de hoogte van het recidiverisico van de jongere. Zo is bijvoorbeeld gebleken dat de effecten voor vaardigheidsprogramma's kleiner zijn bij een uitgebreide geschiedenis van agressie en daardoor een groter recidiverisico (Lipsey, 2009). Ook zijn vaardigheidsprogramma's voor criminele jongeren in een ambulante setting effectiever dan in een residentiële setting in het verminderen van recidive (Lipsey, 2009). Sociale-vaardigheidstrainingen kunnen dus het beste aangeboden worden in een ambulante setting en te agressieve jongeren hebben onvoldoende baat bij de interventie.

² Zoals gedragsscores van observatie/leerkracht/ouder, verwerven van vaardigheden, zelfrapportage van bijvoorbeeld zelfvertrouwen en sociaal aanpassingsvermogen waaronder recidive (Ang & Hughes, 2002).

Ten tweede moeten interventies voldoen aan het *behoeftebeginsel* door zich te richten op factoren waarvan bekend is dat die samenhangen met recidive. Het verminderen van risicofactoren voor delinquentie zou moeten leiden tot een vermindering in recidive (Andrews & Dowden, 2007; Andrews & Bonta, 2010b). Sociale-vaardigheidstrainingen voldoen daaraan doordat zij zich richten op sociale-vaardigheidstekorten. Daarvan is bekend dat die samenhangen met criminaliteit (Dishion, Loeber, Stouthamer-Loeber & Patterson, 1984; Freedman, Rosenthal, Donahoe Jr., Schlundt & McFall, 1978; Gaffney & McFall, 1981; Laak et al., 2003; Larson, Whitton, Hauser & Allen, 2007). Het verband tussen verbeteringen in dynamische risicofactoren (zoals sociale vaardigheidstekorten) en een daarop volgende vermindering in recidive is echter in empirisch onderzoek nog maar zelden aangetoond (Andrews & Bonta, 2010a; Bonta & Andrews, 2007). Voor sociale-vaardigheidstrainingen kan dit samenhangen met het feit dat de effecten op sociale en cognitieve vaardigheden op de langere termijn beperkt zijn (Ang & Hughes, 2002; Cook et al., 2008; Maag, 2006), waardoor een uitwerking op recidive beperkt kan zijn. Bovendien vonden onderzoeken naar delinquente jongeren in Nederland die positieve effecten op dynamische risicofactoren (vaardigheden) meteen na de interventie lieten zien, tot nu toe geen lange termijn vermindering van recidive (MST, Asscher et al., 2014; EQUIP, Brugman & Bink, 2011; Helmond, Overbeek & Brugman, 2015).

Ten derde moeten interventies voldoen aan het *responsiviteitsbeginsel* dat stelt dat de interventie een *match* moet zijn tussen dader, uitvoerder en programma en methodiek. Voor sociale-vaardigheidstrainingen betekent dit dat jongeren geïndiceerd moeten worden die de (cognitieve) mogelijkheid hebben om te profiteren van de technieken die in de training worden ingezet (Cook et al., 2008).

Ten vierde moet voldaan worden aan het beginsel van *behandelmodaliteit*. Dit beginsel stelt dat de interventie op meerdere gebieden moet aangrijpen en met meerdere methodieken moet werken. In bestaand onderzoek naar sociale-vaardigheidstrainingen bleek het generaliseren van de geleerde vaardigheden in de training naar daarbuiten een probleem (Cook et al., 2008; Maag, 2006), wat mogelijk leidde tot verminderde effecten op de lange termijn.

Ten vijfde moet voldaan worden aan het beginsel van *programma-integriteit*. Dit houdt in dat de interventie in de praktijk wordt uitgevoerd, zoals die is bedacht of omschreven. In eerder onderzoek naar sociale-vaardigheidstrainingen bleek dat het vaak aan (aantoonbare) programma-integriteit ontbrak (Cook et al., 2008; Gresham, Cook, Crews, & Kern, 2004; Maag, 2006), terwijl vaardigheidsprogramma's met een hoge behandelintegriteit het effectiefst waren in het verminderen van recidive (Goense, Assink, Stams, Boendermaker & Hoes, 2016; Lipsey, 2009).

Tot slot moet voldaan worden aan het *professionaliteitsbeginsel*, wat stelt dat de trainers goed opgeleid en begeleid moeten zijn. Tot nu toe werd in onderzoek naar sociale-vaardigheidstrainingen de professionaliteit van de trainer nog amper gedefinieerd (Cook et al., 2008) waardoor aan dit principe nog onvoldoende voldaan werd.

1.3 Onderzoek naar Tools4U tot nu toe

In tegenstelling tot veel van de interventies die in eerdergenoemde overzichtsstudies zijn meegenomen, voldoet Tools4U aan alle kenmerken waarvan bekend is dat die bijdragen aan de effectiviteit van interventies voor criminele jongeren (*What Works*-beginselen, Andrews & Bonta, 2010a; Bonta & Andrews, 2007; Van der Laan, 2004). Dit is de afgelopen jaren niet alleen op basis van de theoretische onderbouwing aangetoond, maar ook aan de hand van empirisch onderzoek in Nederland.

Het eerste onderzoek was een procesevaluatie, waaruit bleek dat er knelpunten waren in zowel de indicering als de uitvoering van Tools4U (Bosma et al., 2011). Nadat er verschillende verbeteringen in het indiceringsproces waren doorgevoerd, werd een procesevaluatie en doeltreffendheidsonderzoek uitgevoerd (Van der Stouwe et al., 2013). Uit dit vragenlijstonderzoek onder 115 jongeren en een dossiersteekproef bij de Raad van de Kinderbescherming (RvdK) van jongeren die landelijk Tools4U kregen, bleek dat de jongeren die in de praktijk de interventie kregen grotendeels voldeden aan de indicatiecriteria. Door de ambulante setting, het indiceren van jongeren met een niet al te hoog recidiverisico voor deze relatief lichte interventie en het uitsluiten van bijvoorbeeld jongeren met agressieproblematiek voldoet de interventie dus aan het *risicobeginsel*. Er waren echter ook enkele jongeren (17%) geïnccludeerd die geen delict hadden gepleegd, maar de training opgelegd kregen voor spijbelen. In aansluiting op het *risicobeginsel* zou de training voor deze jongeren niet gepast zijn, omdat er geen sprake kan zijn van een recidiverisico, wanneer er (nog) geen delict gepleegd is.

Daarnaast bleek dat de training in de praktijk grotendeels volgens de handleiding uitgevoerd werd (Van der Stouwe et al., 2013). Verschillende onderdelen die aansluiting bij de jongere moeten waarborgen, zoals voldoende motivatie van de jongere, toetsing van de match tussen trainer en de jongere en het gezamenlijk opstellen van een persoonlijk trainingsplan (Albrecht & Spanjaard, 2011) werden in de praktijk voldoende toegepast, waarmee *responsiviteit* aannemelijk lijkt. Daarnaast bleek dat Tools4U voldoende programma-integer werkt, dat alle verplichte trainingsonderdelen en -technieken voldoende toegepast werden en dat de trainers voldoende opgeleid, begeleid en beoordeeld werden, hoewel het lastig bleek om ouders voldoende te betrekken bij de training (Van der Stouwe et al., 2013). Daarmee voldoet de interventie aan zowel het beginsel van *programma-integriteit* als het *professionaliteitsbeginsel* en het beginsel van *behandelmodaliteit*.

Vervolgens werd er een effectstudie uitgevoerd, waarbij de effecten van Tools4U op sociale en cognitieve vaardigheden werden onderzocht (Van der Stouwe et al., 2015; Van der Stouwe, Asscher, Hoeve, Van der Laan, & Stams, 2016). Daarvoor werd de Tools4U-groep uit de doeltreffendheidsstudie vergeleken met een controlegroep van 108 jongeren die ongeveer een jaar later andere werk- of leerstraf kregen. Deze jongeren waren op basis van PSM als meest vergelijkbare jongeren geselecteerd. Uit deze effectstudie bleek dat Tools4U meteen na afloop van de leerstraf effectiever was dan bij een controlegroep in het verminderen van specifieke cognitieve en sociale vaardigheden waarmee een verband met criminaliteit is aangetoond: impulsiviteit (Cruise et al., 2008; Veltri et al., 2014), denkfouten (Helmond, Overbeek, Brugman & Gibbs, 2015; Paternoster, Pogarsky & Zimmerman, 2011), vijandige bedoelingen toekennen (Dodge, Price, Bachorowski & Newman, 1990; Nas, Brugman & Koops, 2005) en adequaat straffen door de ouders (betrokken ouders hebben een beschermende invloed tegen criminaliteit, Hoeve et al., 2009; Lösel & Farrington, 2012). Hoewel daarmee niet aangetoond is dat de interventie insteekt op wat de jongere nodig heeft (en daarmee voldoet aan het *behoeftebeginsel*), tonen deze uitkomsten wel aan dat het Tools4U lukt om de bedoelde vaardigheden daadwerkelijk te verbeteren.

Op basis van deze onderzoeken kan gesteld worden dat Tools4U niet alleen in theorie, maar ook in de praktijk grotendeels voldoet aan de *What Works*-beginselen. Met name het feit dat er voldoende programma-integer gewerkt wordt en gezien het feit dat de risicofactoren waar Tools4U verbeteringen in heeft laten zien belangrijke risicofactoren voor delinquentie zijn (denkfouten, impulsiviteit en het toekennen van vijandige bedoelingen aan anderen, Cruise et al., 2008; Helmond et al., 2015; Lösel & Farrington, 2012; Nas, Orobio De Castro & Koops, 2005; Veltri et

al., 2014), werd verwacht dat de interventie effectief zou zijn in het verminderen van recidive.

Om dit te toetsen werd in 2016 een recidivestudie uitgevoerd (Van der Stouwe et al., 2017). Voor deze studie werd gebruikgemaakt van de steekproef uit het eerdere onderzoek naar de effecten van Tools4U op sociale en cognitieve vaardigheden (Van der Stouwe et al., 2015; Van der Stouwe et al., 2016). De recidivedata werden opgevraagd bij de Justitiële Informatiedienst (JustId) en zowel overtredingen als misdrijven werden meegenomen in de analyses. In dit onderzoek werd geen significant verschil in recidive gevonden tussen de Tools4U-deelnemers en de controlegroep na 6 maanden, 12 maanden, 1,46 jaar³ en de maximaal beschikbare observatieduur⁴ (Van der Stouwe et al., 2017). Er werden bovendien geen modererende effecten gevonden voor leeftijd, geslacht, herkomst en delict geschiedenis. Er werd wel een marginaal significant verschil gevonden in geweldsdelicten: als Tools4U-jongeren recidiveerden, dan deden zij dit minder vaak met een geweldsdelict dan jongeren in de controlegroep.

1.4 Recidive onderzoek

In het huidige onderzoek wordt opnieuw een recidivemeting gedaan, met een recidivetermijn van drie jaar, om te bepalen hoe bestendig de effecten van Tools4U op langere termijn zijn. De doelstelling is om te bepalen in welke mate en op welke wijze Tools4U bijdraagt aan het voorkomen van recidive. In de vorige studie was de recidive voor zowel de Tools4U-groep als de vergelijkingsgroep relatief laag (Van der Stouwe et al., 2017), wat te verwachten was op basis van de geïndiceerde jongeren met een niet al te hoog recidiverisico. Het is daardoor echter mogelijk dat (kleine) significante verschillen pas op langere termijn zichtbaar worden. De data zijn nu opgevraagd uit de OBJD, waarin de 'ruwe' data van het Justitieel Documentatiesysteem (JDS) zijn opgeschoond en gegroepeerd naar dader, waardoor deze data beter vergeleken kunnen worden met gegevens over andere daders en interventies. Daarnaast biedt de huidige meting de mogelijkheid om te onderzoeken in hoeverre het 'nadruppeleffect' (ontstaan door een langere observatieduur voor de Tools4U-groep vergeleken met de controlegroep) een rol heeft gespeeld bij de eerste recidivemeting. Hoewel dit effect niet uit te sluiten is, wordt het over tijd wel kleiner.

Omdat de eerste recidivestudie geen significante effecten vond voor Tools4U, is van belang goed uit te zoeken waar dat aan zou kunnen liggen. Herhaaldelijk hebben onderzoekers aangetoond dat effectiviteit niet bekeken zou moeten worden voor de gehele (heterogene) deelnemersgroep, maar voor specifieke doelgroepen (Kazdin, 2007; Kazdin, 2008; Kraemer, Frank & Kupfer, 2006; Kraemer, Wilson, Fairburn & Agras, 2002). Het is namelijk nog onduidelijk bij welke jongeren, met welke eigenschappen, Tools4U het meeste effect heeft en bij welke jongeren het minste. Wanneer dat bekend is, kan beter aan de eerste drie *What Works*-principes (RNR-principes, Andrews, Bonta & Hoge, 1990; Andrews & Dowden, 2007; Andrews & Bonta, 2010b) worden voldaan. Jongeren met specifieke kenmerken kunnen dan wel of juist niet toegeleid worden naar de interventie. Daarom wordt in de huidige studie een aanvullende moderatoranalyse uitgevoerd om te bepalen voor wie in het bijzonder de interventie effect heeft. Aangezien het eerdere recidiveonderzoek liet zien dat Tools4U niet meer of minder effectief is voor jongeren met bepaalde statische ach-

³ De langst mogelijke observatietermijn die voor alle jongeren beschikbaar was.

⁴ Gemiddeld: Totaal = 2,76 jaar; Tools4U = 3,34 jaar, Controlegroep = 2,13 jaar

tergrondkenmerken (leeftijd, geslacht, herkomst of delictgeschiedenis) (Van der Stouwe et al., 2017), wordt nu ook onderzocht in hoeverre dynamische risicofactoren (vaardigheidstekorten) bij aanvang van de training een rol spelen bij de effectiviteit van Tools4U. Omdat Tools4U zich specifiek op deze vaardigheden richt, is het van belang om te onderzoeken of het niveau van die vaardigheden van invloed is om baat te hebben bij de technieken die in de training aangeboden worden. De uitkomsten van deze moderatoranalyses kunnen helpen om in de toekomst de meest geschikte jongeren voor Tools4U te indiceren, beter te voldoen aan de *What Works*-beginselen en daarmee effectiever te zijn, dan wel om Tools4U aan te passen voor die jongeren voor wie dat nodig is om effectiever te zijn.

De volgende onderzoeksvragen worden in het huidige rapport beantwoord:

- In welke mate is Tools4U effectief in het voorkomen van recidive? Op basis van:
 - prevalentie;
 - ernst eerste recidivedelict.
- Voor welke jongeren is Tools4U meer of minder effectief? Op basis van:
 - leeftijd, geslacht, herkomst, delict geschiedenis;
 - vaardigheidstekorten voorafgaand aan de training.

In het eerdere effectonderzoek zijn de Tools4U-jongeren aan jongeren uit de controlegroep gematcht op basis van een veelvoud aan informatie die bij de voormeting was verzameld: geslacht, leeftijd, herkomst, behandelmotivatie, denkfouten, empathie, impulsiviteit, copingstijlen en sociale-informatieverwerking (Van der Stouwe et al., 2015). Daarmee werd gewaarborgd dat de Tools4U en de controlegroep wat betreft de uitkomstmaten het meeste op elkaar leken. Voor deze methode dienden echter wel landelijk 469 jongeren een vragenlijst in te vullen: een tijdrovende en kostbare methode. Praktischer zou het zijn om een matching te doen op basis van informatie die beschikbaar is in registratiedata. Relevante kenmerken die van daaruit gebruikt kunnen worden ten behoeve van de matching, zijn persoonskenmerken (zoals geslacht, leeftijd en geboorteland), de delictsgeschiedenis en delictskennmerken (zoals type, ernst en opgelegde straf). Omdat tot nu toe nog niet duidelijk is hoe betrouwbaar een dergelijke matching is, wordt in de huidige studie onderzocht of een methode waarbij alleen de OBJD als bron voor matching wordt gebruikt een alternatief kan bieden.

2 Methode

Dit hoofdstuk gaat in op de methode van onderzoek. In paragraaf 2.1 wordt toegelicht welke gegevensbronnen zijn gebruikt. Vervolgens wordt in paragraaf 2.2 beschreven hoe de onderzoeksgroepen zijn samengesteld. In paragraaf 2.3 volgt een bespreking van de gemeten uitkomstmaten, waarna in paragraaf 2.4 wordt ingegaan op de statistische analyses die zijn uitgevoerd.

2.1 Gebruikte gegevensbronnen

In dit onderzoek zijn gegevens uit verschillende bronnen met elkaar gecombineerd. De primaire gegevensbron is een bestand samengesteld voor onderzoek naar de effecten van Tools4U op cognitieve en sociale vaardigheden (Van der Stouwe et al., 2015). Dit bestand bevat gegevens van 115 jeugdige justitiabelen die in 2012 een Tools4U-training hebben doorlopen en van 108 jeugdige justitiabelen die in 2013 of 2014 een andere leer- of werkstraf kregen opgelegd. Van al deze jeugdige justitiabelen zijn, middels diverse vragenlijsten en dossieranalyse van de dossiers van de RvdK, gegevens verzameld over demografische kenmerken, de inhoud van de taakstraf, motivatie voor behandeling en sociale en cognitieve vaardigheden (onder andere denkfouten, empathie, impulsiviteit en coping). Verder bevat het bestand (versleutelde) parketnummers waarmee op individueel niveau een koppeling kan worden gemaakt met data uit de OBJD. De OBJD is een 'versleutelde' versie van het JDS en is bedoeld voor wetenschappelijk onderzoek.⁵ In de OBJD staat de complete justitiële geschiedenis geregistreerd van personen die minimaal 12 jaar oud zijn en die in Nederland verdachte waren in een strafzaak. Het gaat dus alleen om de criminaliteit die onder de aandacht komt van het Openbaar Ministerie (OM). Met behulp van de gegevens in de OBJD kan per jeugdige het strafrechtelijke verleden in kaart worden gebracht. Daarnaast vormt de OBJD de basis voor de recidivemetingen. In dit onderzoek wordt de OBJD tevens gebruikt als basisbestand waaruit personen voor de tweede controlegroep worden geselecteerd. Hiervoor is aanvullende informatie gebruikt van de RvdK over de einddata van de taakstraffen die bij de RvdK zijn uitgevoerd in 2012.

2.2 Samenstelling onderzoeksgroepen

In dit onderzoek worden drie onderzoeksgroepen meegenomen: de Tools4U-deelnemersgroep en twee controlegroepen.

Tools4U-deelnemersgroep

Voor de Tools4U-deelnemersgroep zijn jongeren benaderd die in de periode van 31 mei 2012 tot en met 7 augustus 2012 zijn gestart met een training (Van der Stouwe et al., 2013). De gegevensverzameling werd beëindigd op het moment dat het vooraf vastgestelde aantal van 115 jongeren behaald was. Van deze jeugdigen zijn er tien vroegtijdig gestopt met de training. In de raadsdossiers stond een parketnummer genoteerd naar aanleiding waarvan Tools4U was opgelegd. In de OBJD is ver-

⁵ Alle persoonlijke gegevens (zoals bijvoorbeeld parketnummers of jds-nummers) zijn middels een versleutelingsprocedure veranderd.

volgens nagegaan of in deze zaken daadwerkelijk een leerstraf was opgelegd. Dit was op negen zaken na altijd het geval. In de negen gevallen waar geen leerstraf stond geregistreerd in de OBJD, is gekeken naar andere zaken die in dezelfde periode zijn afgedaan en waarbij wel een leerstraf is opgelegd. Bij vijf personen is uiteindelijk een andere zaak als uitgangzaak gekozen, omdat waarschijnlijker is dat in die zaak de Tools4U-training is opgelegd.⁶ Het startpunt van de recidivemeting van de Tools4U-deelnemers is de datum waarop de Tools4U-training is afgelopen. Daarbij is de datum van de nameting aangehouden als einddatum van de training. In het geval geen nameting gedaan was, is de datum van uitschrijving uit het dossier van de RvdK aangehouden. De dataverzameling uit de OBJD is op 7 juli 2016 afgesloten. Dat betekent dat de observatietermijn van de Tools4U-deelnemers gemiddeld 4,7 jaar bedraagt.

Controlegroep 1

Voor de samenstelling van Controlegroep 1 hebben Van der Stouwe en collega's (2015) gegevens verzameld van 354 jeugdige justitiabelen die in 2013 of 2014 een werkstraf, leerstraf of een combinatie kregen opgelegd. Uit deze groep hebben zij met behulp van PSM op diverse achtergrondkenmerken en de voormeting schalen van de sociale en cognitieve vaardigheden Controlegroep 1 geselecteerd. In de eerste kolom van tabel 1 staat een overzicht van alle kenmerken waarop Controlegroep 1 is gematcht aan de Tools4U-deelnemers. In de uiteindelijke controlegroep zijn 108 jeugdigen opgenomen. Dit betekent dat het voor zeven deelnemers niet mogelijk was om een goede match te vinden. Na de matching werden er echter nauwelijks verschillen tussen de onderzoeksgroepen gevonden en dus wordt aangenomen dat de matching geslaagd is (Van der Stouwe et al, 2015). In hoofdstuk 3 wordt nagegaan of Controlegroep 1 en de Tools4U-deelnemers ook niet verschillen op kenmerken uit de OBJD. Van één persoon uit Controlegroep 1 was te weinig informatie bekend om deze persoon op te kunnen zoeken in de OBJD. Deze persoon kan daarom niet verder worden meegenomen in het recidiveonderzoek. Ook bij de uitgangszaken van personen in Controlegroep 1 zijn we nagegaan of in de OBJD daadwerkelijk een leer- of werkstraf is opgelegd. Dat was in zes zaken niet het geval. Van vijf personen was het mogelijk om in dezelfde periode een andere zaak als uitgangzaak te noteren waarbij wel een werkstraf opgelegd was.⁷

De begindatum van de recidivemeting is de datum waarop de leer- of werkstraf is afgerond. Hiervoor is ofwel de datum van de nameting of de uitschrijvingsdatum van de RvdK aangehouden, afhankelijk van welke eerder was. Het was namelijk mogelijk dat de nameting weken later plaatsvond dan het moment dat de straf was afgerond. Het was echter ook mogelijk dat de RvdK een paar weken na de afronding van de straf de jeugdige pas uitschreef. Vandaar dat ervoor is gekozen om de vroegste datum aan te houden. De observatietermijn van de jeugdigen in Controlegroep 1 is gemiddeld 3,5 jaar.

Controlegroep 2

Voor de tweede controlegroep is een selectie gemaakt van alle jeugdige daders die in 2012 een werkstraf hebben afgerond. Die werkstraf kan zijn opgelegd door het OM of de kinderrechter. Personen die naast de werkstraf ook een leerstraf kregen opgelegd zijn uit de potentiële controlegroep verwijderd, omdat niet uitgesloten kan

⁶ De vier Tools4U-deelnemers waarbij in de OBJD geen leerstraf geregistreerd stond, zijn wel meegenomen in de analyses omdat zeker is dat zij aan Tools4U hebben deelgenomen.

⁷ De jeugdige waarbij geen werkstraf in de OBJD geregistreerd stond, is wel in de verdere analyses meegenomen omdat zeker is dat die persoon wel met de werkstraf gestart is.

worden dat die leerstraf mogelijk een Tools4U-training betreft. Voor matching bestaat Controlegroep 2 uit 9.608 personen. Ook voor de selectie van Controlegroep 2 is gebruikgemaakt van PSM. In de tweede kolom van tabel 1 is opgesomd op welke achtergrondkenmerken Controlegroep 2 vergelijkbaar is gemaakt met de Tools4U-deelnemersgroep. Uiteindelijk bestaat Controlegroep 2 uit 115 personen. In hoofdstuk 3 zal nagegaan worden in hoeverre de matching geslaagd is. De recidivemeting van Controlegroep 2 is gestart op het moment waarop de taakstraf is afgerond, inclusief eindrapportage en administratieve afronding bij de RvdK. De gemiddelde observatietermijn van Controlegroep 2 bedraagt vijf jaar.

Tabel 1 Variabelen waarmee Controlegroep 1 en 2 zijn gematched aan Tools4U-deelnemersgroep

Controlegroep 1	Controlegroep 2
Demografische kenmerken	Demografische kenmerken
Sekse	Sekse
Leeftijd bij start training	Leeftijd bij start training
Herkomst	Geboorteland
	Kenmerken van de uitgangszaak
	Delictcategorie
	Arrondissement
	Modaliteit (OM/kinderrechter)
	Ernst uitgangszaak (MSI-score)
Gegevens strafrechtelijke carrière	Gegevens strafrechtelijke carrière
Zelfgerapporteerde delinquentie	Aantal eerdere strafzaken
	Leeftijd 1e strafzaak
	Strafzaakdichtheid ⁸
Vaardigheden	
Motivatie voor behandeling	
Cognitieve vertekeningen	
Empathie	
Coping stijlen	
Sociale-informatieverwerking	

Status van beide controlegroepen

In het vergelijkend recidiveonderzoek zal het effect van Tools4U gemeten worden aan de hand van een vergelijking tussen de deelnemersgroep en de twee controlegroepen. De twee controlegroepen verschillen op een aantal aspecten van elkaar. Een eerste verschil is dat Controlegroep 2 alleen uit jeugdige daders met een werkstraf bestaat, terwijl in Controlegroep 1 ook jongeren zitten die een leerstraf kregen opgelegd.⁹ Er zijn echter slechts zes personen met een leerstraf opgenomen in Controlegroep 1. We denken daarom dat dit verschil geen substantiële invloed op de resultaten zal hebben.

⁸ De strafzaakdichtheid is het gemiddelde aantal strafzaken dat men had in de periode tussen het jaar van het eerste justitiecontact en dat van de uitgangszaak

⁹ Het doel bij het samenstellen van beide controlegroepen was het selecteren van een groep jeugdigen die zo goed mogelijk vergelijkbaar is met de Tools4U-deelnemers. Je hebt de meeste kans om een goed vergelijkbare groep te vinden als je zowel kunt kiezen uit jeugdigen met een werkstraf en/of een leerstraf (anders dan Tools4U). Vandaar dat in Controlegroep 1 ook jeugdigen met een leerstraf zijn opgenomen. Voor Controlegroep 2 was dit niet mogelijk en zijn alleen jeugdigen met een werkstraf geselecteerd.

Een tweede verschil is dat de recidivemeting van de controlegroepen in een andere periode plaatsvindt. Controlegroep 1 is een jaar later geworven dan de Tools4U-deelnemers terwijl Controlegroep 2 bestaat uit jeugdigen die in dezelfde periode als de Tools4U-deelnemers hun straf hebben afgerond. Dit heeft een aantal mogelijke nadelen. Ten eerste kan er sprake zijn van een zogenoemd 'periode-effect'. Dat wil zeggen dat de recidive in 2012 in het algemeen hoger of lager zou kunnen liggen dan in 2013 en 2014. Het gaat hier echter maar om een verschil van één à twee jaar en daarom is de verwachting dat een eventueel periode-effect beperkt zal zijn. Daarnaast kan de kortere observatietermijn van Controlegroep 1 leiden tot een onderschatting van de recidive aan het einde van de in dit onderzoek gehanteerde recidivetermijn.¹⁰ Dit komt doordat het soms enige tijd duurt voordat nieuwe strafzaken worden geregistreerd in de justitiële documentatie. De resultaten van Controlegroep 1 aan het einde van de survivalcurve zullen dus met enige voorzichtigheid moeten worden geïnterpreteerd.

Het grootste verschil tussen de controlegroepen zit hem echter in de beschikbare informatie om te matchen. Uit tabel 1 wordt duidelijk dat Controlegroep 1 op andere achtergrondkenmerken gematcht is aan de deelnemersgroep dan Controlegroep 2. Zo konden bij de matching van Controlegroep 2 alleen statische kenmerken¹¹ worden betrokken, terwijl bij de matching van Controlegroep 1 ook dynamische factoren zijn gebruikt. Door beide controlegroepen in dit onderzoek te betrekken kan gekeken worden of matching op andere achtergrondkenmerken tot verschillende resultaten leidt.

2.3 Recidivemetingen

De recidivemetingen voor het huidige onderzoek zijn uitgevoerd volgens werkwijze van de WODC-Recidivemonitor (Wartna, Blom & Tollenaar, 2011). Dit betekent dat als recidivedelict zaken worden geteld die zijn afgedaan door de rechter of het OM, maar die niet zijn geëindigd in een vrijspraak, technisch sepot of andere technische beslissing. Zaken die nog niet zijn afgedaan en zaken die in hoger beroep dienen, worden wel meegeteld. De grote meerderheid van de strafrechtelijke vervolgingen eindigt namelijk in een geldige afdoening (De Heer-de Lange & Kalidien, 2014). Verder is het van belang om op te merken dat overtredingen buiten beschouwing worden gelaten. Hier is voor gekozen, omdat in de OBJD slechts een selectie van de overtredingen is opgenomen, de overtredingen waarbij een geldboete is opgelegd van minder dan € 100 ontbreken namelijk (Justitiële Informatiedienst, 2009). In het huidige onderzoek wordt gekeken naar twee facetten van recidive, namelijk de recidiveprevalentie en de ernst van het eerste recidivedelict. De recidiveprevalentie heeft betrekking op het aantal personen dat binnen een bepaalde periode na beëindiging van de sanctie opnieuw met justitie in aanraking kwam wegens een misdrijf. Hierbij wordt ook duidelijk hoe lang het duurt voordat de jeugdigen een delict plegen waarvoor zij opnieuw veroordeeld worden. Daarnaast wordt gekeken naar de ernst van het eerste recidivedelict.¹² Het is immers van belang om, naast de vraag of men heeft gerecidiveerd, de ernst van het nieuwe delict in ogenschouw te

¹⁰ In het huidige onderzoek kijken we bij alle onderzoeksgroepen naar recidivedelicten in een periode van drie jaar na afronding van de leerstraf of werkstraf.

¹¹ Statische kenmerken zijn vaste waarden die niet veranderen of slechts in één richting (Vogelvang, Van Burik, Van der Knaap & Wartna, 2003).

¹² Er is gekozen om alleen de ernst van het eerste recidivedelict te vergelijken omdat er dan geen rekening hoeft te worden gehouden met eventuele vervolgdetenties na het eerste delict.

nemen. De impact van de recidive mag groter worden geacht als een zwaarder recidivedelict is gepleegd. Om een inschatting te maken van de ernst van een strafzaak maken we gebruik van de 'misdrijf-straf index' (MSI; Beerthuizen, Wartna, Verweij & Tollenaar, 2015). De score op deze maat wordt afgeleid van de afdoening van de zaak. De opgelegde straffen, bijvoorbeeld het aantal dagen gevangenisstraf, de boete van een bepaalde hoogte of werkstraf met een bepaalde duur, worden omgezet naar een bepaald aantal 'sanctiepunten' of detentiedageequivalenten (dde).¹³ In dit onderzoek moet rekening worden gehouden met het gegeven dat een persoon in een eventuele recidivezaak als jeugdige of als volwassene zou kunnen worden berecht. De MSI-scores van jeugdzaken zullen in het algemeen lager liggen, omdat binnen het jeugdstrafrecht minder zware straffen kunnen worden opgelegd. Het is echter niet zo dat jeugdigen per definitie minder ernstige delicten plegen. Zodra de ernst van recidivezaken van verschillende personen wordt vergeleken die niet volgens hetzelfde strafrecht zijn berecht, kunnen er dus vertekeningen optreden. Van daar dat in dit onderzoek de MSI-scores van volwassenzaken zijn omgezet naar vergelijkbare scores van jeugdzaken (voor de methode, zie Verweij et al., 2017). Elke strafzaak krijgt uiteindelijk een aantal sanctiepunten variërend van 0,1 tot 100.

2.4 Statistische analyses

Voor de analyse van de gegevens hebben we gebruikgemaakt van verschillende analysetechnieken. In deze paragraaf worden deze kort toegelicht. In bijlage 2 wordt op de meer technische details van de gebruikte analysetechnieken ingegaan.

Propensity score matching (PSM)

Voor de matching van Controlegroep 2 is gebruikgemaakt van PSM (Rosenbaum & Rubin, 1983).¹⁴ Een *propensity score* is een met behulp van regressieanalyse op de gemeten achtergrondkenmerken berekende waarschijnlijkheid dat een persoon aan een bepaalde conditie wordt toegewezen. In dit onderzoek is dat de kans dat men een Tools4U-training kreeg opgelegd. Een persoon in de Tools4U-deelnemersgroep wordt vergelijkbaar geacht met iemand in de controlegroep als zij dezelfde *propensity score* hebben. Een gelijke *propensity score* houdt in dat voor een paar gemaakte personen, men aanneemt dat de toewijzing aan de conditie volledig willekeurig is, gegeven de gelijke *propensity score*. Voor de controlegroep wordt de persoon geselecteerd van wie de *propensity score* het dichtst ligt bij die van de persoon in de behandelgroep.¹⁵ Een persoon kan niet meerdere keren aan een ander gemaakt worden. Als er twee of meer even goede matches zijn, dan vond de selectie van de match willekeurig plaats. Om te checken of de PSM-procedure goed heeft gewerkt, werden de groepen voor en na matching op alle achtergrondkenmerken met elkaar vergeleken.

Survivalanalyse

Om het verloop van de prevalentie van recidive over de tijd te schatten is gebruikgemaakt van survivalanalyse (Kaplan & Meier, 1958). Dit is de standaardmethode om de recidivekans te schatten bij een dataset die gedeeltelijk *gensored* is en waarbij de tijd continue gemeten is. *Censoring* wil zeggen dat de observatie van een

¹³ De basisformule voor de berekening van de dde-score is: 1 dde = € 36 boete = 2 uur taakstraf = 1 celdag.

¹⁴ Voor meer informatie over de matching van Controlegroep 1, zie Van der Stouwe et al. (2015).

¹⁵ In hoeverre het lukt om personen met vergelijkbare *propensity scores* te selecteren, wordt duidelijk in hoofdstuk 3.

persoon stopt. Dat kan zijn omdat diegene recidiveert of omdat de gegevensverzameling wordt beëindigd. Aangezien niet iedereen even lang gevolgd kan worden, treden er verschillen op in observatieduur. Door middel van survivalanalyse kan desondanks een betrouwbare schatting van de recidivekans worden verkregen.

Multipele survival regressieanalyse en penalized survival regressie

Ten behoeve van de beantwoording van de derde onderzoeksvraag (voor welke jongeren is Tools4U meer of minder effectief?) is een multipele survival regressieanalyse uitgevoerd. Een multipele survival regressieanalyse maakt het mogelijk een indicatie te krijgen van de unieke invloed van verschillende kenmerken op de recidive. In het huidige onderzoek maken we gebruik van een parametrisch survival regressiemodel, zie bijlage 2 voor meer informatie over deze analysemethode. Voor de multipele survival regressieanalyse, waarin we ook willen kijken naar interactie-effecten tussen deelname aan Tools4U en achtergrondkenmerken en vaardigheids-tekorten, is een groot aantal variabelen beschikbaar. Omdat de onderzoeksgroep vrij klein is ontstaat het gevaar dat het model te complex wordt. In dat geval is de geschatte invloed van de achtergrondkenmerken op recidive niet meer betrouwbaar. Het is daarom van belang om het aantal variabelen niet te groot te maken. Om die reden wordt met behulp van een *penalized* survival regressiemodel gekeken welke set van variabelen de uitkomst het beste voorspelt. Een voordeel van deze methode is dat de selectie van predictoren tegelijkertijd plaatsvindt en dat met deze methode ook meteen wordt bepaald hoeveel predictoren in het model worden opgenomen.

3 Beschrijving onderzoeksgroepen

In dit hoofdstuk worden de onderzoeksgroepen die in dit onderzoek centraal staan beschreven. Het is van belang dat de jeugdigen in de controlegroepen vergelijkbaar zijn met de jeugdigen die Tools4U kregen opgelegd. Alleen als dat het geval is, kunnen de gevonden effecten worden toegeschreven aan de interventie. Vandaar dat in dit hoofdstuk ingaan wordt op de vergelijkbaarheid van de onderzoeksgroepen. Paragraaf 3.1 beschrijft de kenmerken van de Tools4U-deelnemers. In paragraaf 3.2 kijken we naar de vergelijkbaarheid van de Tools4U-deelnemers en Controlegroep 1. Hierbij richten we ons alleen op de verschillen op OBDJ-kenmerken tussen de Tools4U-deelnemers en Controlegroep 1 na matching. Hier is voor gekozen omdat in eerder onderzoek al is getoond dat de Tools4U-deelnemers en Controlegroep 1 vergelijkbaar zijn wat betreft de kenmerken waarop Controlegroep 1 samengesteld is (Van der Stouwe et al., 2015). In paragraaf 3.3 komen de verschillen aan bod tussen de Tools4U-deelnemers en Controlegroep 2. Omdat Controlegroep 2 in dit onderzoek met PSM op basis van de OBDJ-kenmerken wordt samengesteld, gaan we in paragraaf 3.3 zowel in op de verschillen vóór als na matching.

3.1 Achtergrondkenmerken Tools4U-deelnemers

Tabel 2 geeft een overzicht van de persoonskenmerken, de kenmerken van de uitgangszak en de criminele carrière, zoals die geregistreerd staan in de OBDJ.¹⁶ In de kolommen staan achtereenvolgens de kenmerken van de Tools4U-deelnemers, Controlegroep 1 na matching, Controlegroep 2 voor matching en Controlegroep 2 na matching.

De eerste kolom van tabel 2 laat zien dat ongeveer driekwart van de deelnemers bestond uit mannen (73,0%). De deelnemers zijn gemiddeld 16 jaar oud bij de start van Tools4U. Het grootste gedeelte van de deelnemers is in Nederland geboren (89,6%). Het type delict waarvoor de Tools4U-deelnemers zijn veroordeeld is divers. Bijna een derde van de Tools4U-deelnemers is veroordeeld voor een vermogensdelict (30,4%) en meer dan 10% van de deelnemers pleegde een geweldsdelict, een vermogensdelict met geweld of een delict dat valt in de categorie vernieling, lichte agressie en openbare orde. In de deelnemersgroep zitten geen deelnemers die een zedendelict hebben gepleegd. Het eerdere effectiviteitsonderzoek van Van der Stouwe en collega's (2015) liet al zien dat niet alle Tools4U-deelnemers een delict hebben gepleegd. Ongeveer 17% neemt deel aan de training vanwege schoolverzuim.¹⁷

De meerderheid van de Tools4U-deelnemers kreeg de straf opgelegd door de kinderrechter. In de deelnemersgroep zitten verder vooral jongeren die veroordeeld werden in de arrondissementen Noord Holland, Amsterdam, Rotterdam en Noord-Nederland. De Tools4U-deelnemers hadden gemiddeld op de leeftijd van 14,9 jaar

¹⁶ De zaaksinformatie zoals geregistreerd in de OBDJ en de dossiers van de RvdK komt niet altijd overeen. Deels komt dit omdat we in een aantal gevallen een andere uitgangszak hebben geselecteerd (zie paragraaf 2.2).

¹⁷ Hoewel in de OBDJ niet alle overtredingen geregistreerd staan (overtredingen die zijn afgedaan met een boete van minder dan € 100 ontbreken), speelt dat hier geen rol omdat de afdoening bij deze onderzoeksgroepen een leer- of werkstraf betreft.

het eerste justitiecontact. Bij de start van Tools4U hebben de deelnemers gemiddeld 0,8 eerdere justitiecontacten.

Tabel 2 OBD-achtergrondkenmerken Tools4U-deelnemers en jeugdigen in de controlegroepen

Tools4U-deelnemers	N=115	Na matching				Voor matching				Na matching			
		Cg 1	d	t/X ²	p	Cg 2	d	t/X ²	p	Cg 2	d	t/X ²	p
Persoonskenmerken													
Sekse (man) %	73,0	71,0	0,04	0,12	75,9	-0,07	0,54		74,8	-0,04	0,09		
Gemiddelde leeftijd	16,0	15,9	0,07	0,55	16,4	-0,26	-2,89	**	15,9	0,04	0,30		
Geboorteland (niet Nederland) %	10,4	14,0	-0,11	0,66	11,7	-0,04	0,19		8,7	0,06	0,20		
Kenmerken uitgangzaak													
Delictcategorie %				14,14	*			53,38	**				1,16
Geweld	16,5	17,8	-0,03			15,6	0,03			20,0	-0,09		
Zeden	0,0	0,0	0,00			0,3	-0,06			0,0	0,00		
Vermogen met geweld	14,8	2,8	0,43			3,0	0,68			11,3	0,10		
Vermogen zonder geweld	30,4	28,0	0,05			31,1	-0,01			28,7	0,04		
Vernieling, lichte agressie en openbare orde	14,8	19,6	-0,13			19,7	-0,12			13,9	0,02		
Overig	7,0	3,7	0,14			10,1	-0,11			7,8	-0,03		
Schoolverzuim	16,5	28,0	-0,28			20,1	-0,09			18,3	-0,05		
Modaliteit (kinderrechter) %	63,5	51,4	0,25	3,31		51,7	0,24	6,28	*	62,6	0,02	0,02	
Arrondissement %				37,98	**			60,03	**				3,29
Amsterdam	15,7	32,7	-0,41			8,3	0,27			19,1	-0,09		
Noord-Holland	19,1	29,0	-0,23			6,2	0,53			17,4	0,05		
Midden-Nederland	6,1	5,6	0,02			9,6	-0,12			4,3	0,08		
Noord-Nederland	13,9	3,7	0,36			11,6	0,07			14,8	-0,02		
Den Haag	6,1	2,8	0,16			14,3	-0,24			7,0	-0,04		
Rotterdam	14,8	0,9	0,52			12,3	0,08			13,0	0,05		
Limburg	9,6	1,9	0,33			6,0	0,15			13,0	-0,11		
Oost-Brabant	1,7	2,8	-0,07			7,9	-0,23			0,9	0,08		
Zeeland-West Brabant	4,3	5,6	-0,06			8,5	-0,15			1,7	0,15		
Oost-Nederland	3,5	7,5	-0,18			9,6	-0,21			3,5	0,00		
Gelderland													
Oost-Nederland	5,2	7,5	-0,09			5,7	-0,02			5,2	0,00		
Overijssel													
Gemiddelde ernst uitgangzaak	41,4	40,5	0,10			38,8	0,26			40,5	0,10		
Criminele carrièrekenmerken													
Gemiddelde leeftijd 1e strafzaak	14,9	15,2	-0,22	-1,66	*	15,5	-0,41	-4,56	**	14,8	0,04	0,27	
Gemiddelde aantal eerdere strafzaken	0,8	0,7	0,07	0,50		0,7	0,06	0,84		0,8	0,06	0,46	
Gemiddelde strafzaak-dichtheid	0,4	0,3	0,11	0,85		0,3	0,16	1,69	*	0,3	0,06	0,42	

** p=<0,01; * p<0,05

Noot: Cg 1 = Controlegroep 1, Cg 2= Controlegroep 2.

3.2 Vergelijkbaarheid Tools4U-deelnemers en Controlegroep 1

Controlegroep 1 is op tal van kenmerken gematcht aan de Tools4U-deelnemers, maar niet op de OBDJ-kenmerken. Vandaar dat hier is nagaan of de Tools4U-groep en Controlegroep 1 ook op die kenmerken vergelijkbaar zijn. Uit Tabel 2 blijkt dat op de persoonskenmerken geen significante verschillen zijn gevonden tussen de Tools4U-deelnemers en Controlegroep 1. Een aantal kenmerken van de uitgangszaken verschillen wel significant. De Tools4U-deelnemers blijken bijvoorbeeld vaker een vermogensdelict met geweld te hebben gepleegd dan de jeugdige daders in Controlegroep 1. Daarentegen zitten in Controlegroep 1 meer jeugdigen die een leer- of werkstraf kregen opgelegd voor het overtreden van de Leerplichtwet. Verder zijn de Tools4U-deelnemers relatief iets vaker door de rechter veroordeeld dan Controlegroep 1. Dit verschil is net niet significant ($\chi^2_{(df=1)}=3,31, p=0,069$). De arrondissementen waar de uitgangszaken zijn behandeld, verschillen ook tussen de Tools4U-deelnemers en Controlegroep 1. De Tools4U-deelnemers zijn relatief vaak veroordeeld in Rotterdam, Limburg en Noord-Nederland, terwijl de zaken van Controlegroep 1 relatief vaker zijn afgedaan in Amsterdam en Noord-Holland. De gemiddelde leeftijd van de eerste strafzaak van de jeugdigen in Controlegroep 1 ligt net iets hoger dan in de Tools4U-deelnemersgroep. Bij het vergelijken van de recidivepercentages tussen Controlegroep 1 en de Tools4U-deelnemers zullen we nagaan of deze verschillen van invloed zijn op de uitkomst.

3.3 Vergelijkbaarheid Tools4U-deelnemers en Controlegroep 2

Voor matching

Uit tabel 2 blijkt dat de Tools4U-deelnemers en Controlegroep 2 voor matching op diverse achtergrondkenmerken verschillen. Vanwege de specifieke selectie-eisen voor deelname aan Tools4U (o.a. cognitieve en sociale vaardigheidstekorten die een rol spelen bij het delict en een matig tot hoog risico op herhaling) is het niet onverwacht dat de Tools4U-deelnemers op sommige achtergrondkenmerken afwijken van de totale populatie jeugdigen die een werkstraf kregen opgelegd. Zo was bij de Tools4U-deelnemers een vermogen met geweldsdelict vaker de aanleiding voor de strafzaak. De werkstrafpopulatie kreeg daarentegen vaker een straf opgelegd wegens schoolverzuim. Verder zijn jeugdigen in de Tools4U-groep gemiddeld jonger bij hun eerste justitiecontact en hebben zij gemiddeld iets meer eerdere justitiecontacten. Ook de strafzaakdichtheid¹⁸ van de Tools4U-deelnemers is gemiddeld net iets hoger dan van de werkstrafpopulatie. Via PSM zal een gemiddeld net iets zwaardere groep uit de werkstrafpopulatie moeten worden geselecteerd om een goed vergelijkbare controlegroep samen te stellen.

Matching

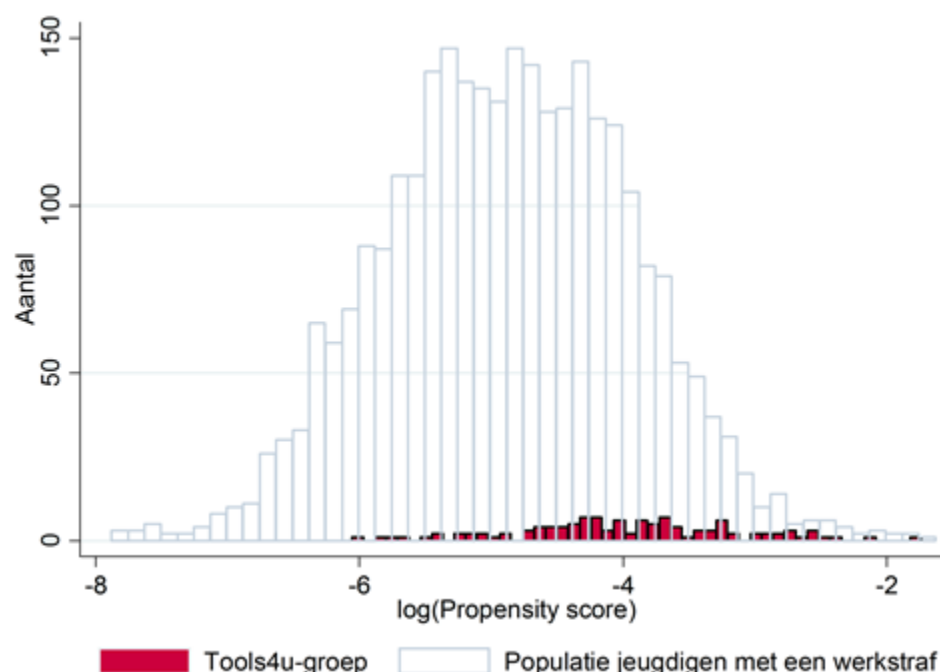
Of het mogelijk lijkt om ondanks de initiële verschillen tussen de Tools4U-deelnemers en de werkstrafpopulatie een goede controlegroep te selecteren, wordt bekeken door vooraf de *propensity scores* van de groepen te vergelijken. Met behulp van regressie is voor iedere persoon in het onderzoek een *propensity score* berekend. Deze score is berekend op basis van de achtergrondkenmerken en staat model voor de kans om Tools4U opgelegd te krijgen. Voor ieder persoon uit de Tools4U-groep is een persoon uit de totale *pool* van jeugdigen met een werkstraf gezocht met een-

¹⁸ De strafzaakdichtheid is het gemiddelde aantal strafzaken dat men had in de periode tussen het jaar van het eerste justitiecontact en dat van de uitgangszaken.

zelfde *propensity score*. Om een goed vergelijkbare controlegroep te krijgen is het dus belangrijk dat de verdeling van de *propensity scores* van de groep Tools4U-deelnemers wordt overlapt door de verdeling van *propensity scores* van de totale werkstrafpopulatie.

In figuur 1 wordt de verdeling van de scores voor de Tools4U-deelnemers (rood) en Controlegroep 2 (doorzichtig wit) afgebeeld. Afgezet tegen Controlegroep 2 hebben de Tools4U-deelnemers relatief middelhoge tot hoge *propensity scores*. De totale werkstrafpopulatie overlapt de groep Tools4U-deelnemers echter volledig. Dit is een indicatie dat het goed mogelijk is om de groepen te matchen op *propensity scores* en ze zo op de beschikbare achtergrondkenmerken vergelijkbaar te maken.

Figuur 1 Overlap (log-)propensity scores Tools4U-deelnemers en populatie jeugdigen met een werkstraf



Noot: voor een betere visualisatie van figuur 1 is random 1/3 van de werkstrafpopulatie geselecteerd en in de figuur weergegeven.

Na matching

Uit de laatste kolom van tabel 2 blijkt dat er na matching slechts kleine verschillen bestaan tussen de Tools4U-deelnemers en Controlegroep 2. Deze verschillen zijn niet significant ($p > 0,05$). Daarom kan geconcludeerd worden dat de Tools4U-deelnemers en de jeugdigen in Controlegroep 2 goed vergelijkbaar zijn wat betreft de kenmerken beschikbaar uit de OBJD.¹⁹

¹⁹ Aangezien Controlegroep 2 vergelijkbaar is gemaakt met de Tools4U-deelnemers is deze controlegroep qua OBJD-achtergrondkenmerken ook goed vergelijkbaar met Controlegroep 1. De verschillen tussen de twee controlegroepen bestaan slechts uit de verschillen op de variabelen waarop Controlegroep 1 met de Tools4U-deelnemers verschilt.

4 Resultaten

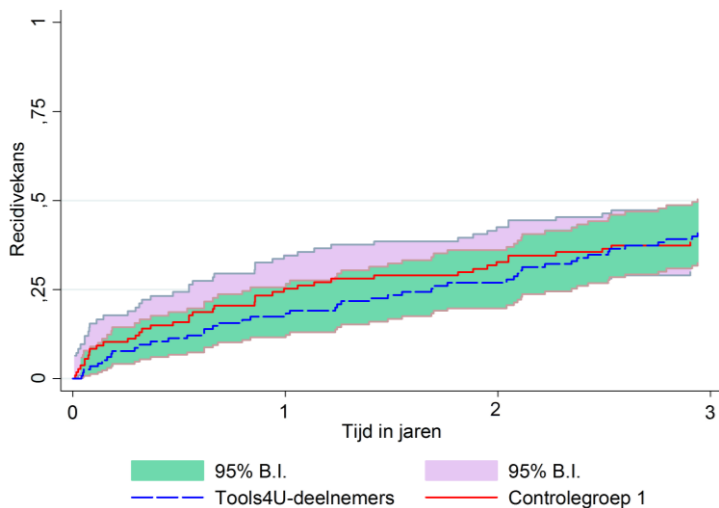
In dit hoofdstuk worden de resultaten van de effectmeting besproken. In paragraaf 4.1 wordt ingegaan op de recidiveprevalentie van de Tools4U-deelnemers en de controlegroepen. Paragraaf 4.2 gaat in op een vergelijking van de ernst van het eerste recidivecontact. Ten slotte wordt in paragraaf 4.3 geprobeerd te achterhalen of deelname aan Tools4U voor specifieke groepen jeugdigen een verschillend effect heeft gehad.

4.1 Recidiveprevalentie

Figuur 2 geeft de ontwikkeling van de recidiveprevalentie in de tijd weer van Tools4U-deelnemers en Controlegroep 1. Hierbij is gekeken tot maximaal drie jaar na afloop van Tools4U of de werk-/leerstraf van de jeugdige in de controlegroep. Van de Tools4U-deelnemers recidiveert 17% binnen één jaar. Deze deelnemers hebben dus binnen een jaar na het afronden van de gedragsinterventie opnieuw een misdrijf gepleegd waarvoor ze in contact komen met justitie. Na twee en drie jaar is dit respectievelijk 27% en 41%. De één- en tweejarige recidivepercentages van de Controlegroep 1 liggen hoger. Na één jaar recidiveert 25% van de jeugdigen in Controlegroep 1 en de tweejarige recidiveprevalentie bedraagt 33%. Het driejarige recidivepercentage van Controlegroep 1 ligt met 38% net iets lager dan de recidiveprevalentie van de Tools4U-deelnemers. Er zijn drie toetsen gedaan (de Log-ranktoets, de Wilcoxon-toets en de Tarone-Waretoets), die elk op een andere fase van de curve afgebeeld in de figuur de nadruk legt, om de verschillen in recidive te toetsen. Uit geen van deze toetsen volgt dat het verschil in recidiveprevalentie significant is. Dit betekent dat zowel aan het begin van de curve als aan het einde van de drie jaar geen statistisch significante verschillen gevonden zijn. We concluderen dan ook dat wat betreft de prevalentie van recidive geen statistisch significante verschillen gevonden zijn tussen de Tools4U-deelnemers en Controlegroep 1.²⁰

²⁰ De Tools4U-deelnemers en Controlegroep 1 verschillen licht op het type delict van de uitgangszaak, de gemiddelde leeftijd van de eerste strafzaak, de instantie die de uitgangszaak heeft behandeld en het arrondissement waar de uitgangszaak is behandeld (zie hoofdstuk 3). Indien met deze kenmerken rekening wordt gehouden in een survival regressie model dan wordt nog steeds een niet significant effect van de Tools4U-training gevonden. We concluderen daarom dat de verschillen geen invloed lijken te hebben op de uitkomst.

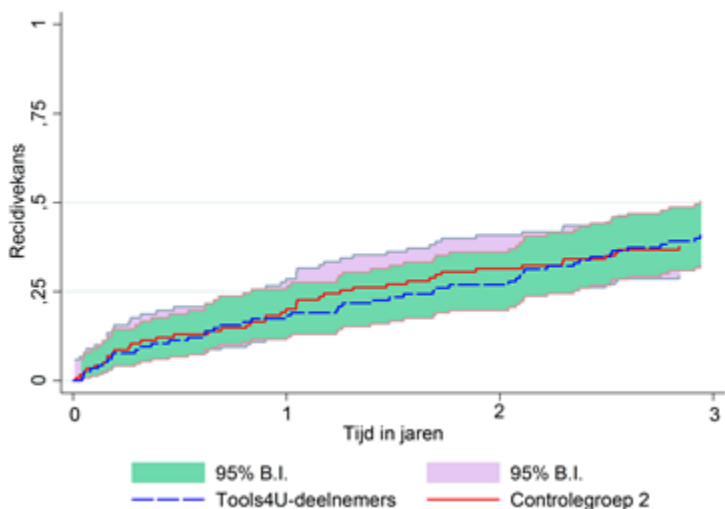
Figuur 2 Recidiveprevalentie Tools4U-deelnemers vergeleken met Controlegroep 1



Noot: Log-rank = 0,01, $p=0,92$; Wilcoxon = 0,10, $p=0,75$; Tarone-Ware = 0,04, $p=0,84$.

De recidiveprevalentie van de Tools4U-deelnemers is ook vergeleken met die van Controlegroep 2 (zie figuur 3). In het eerste jaar recidiveert 20% van de jeugdigen in Controlegroep 2 en in het tweede jaar 32%. De jeugdigen in Controlegroep 2 hebben dus in de eerste twee jaar meer gerecidiveerd dan de Tools4U-deelnemers. Na drie jaar is dit omgekeerd, dan recidiveert 38% van Controlegroep 2 terwijl het recidivepercentage van de Tools4U-deelnemers 41% bedraagt. De recidive-curves van de Controlegroep 2 en de Tools4U-deelnemers verschillen echter ook niet significant van elkaar.

Figuur 3 Recidiveprevalentie Tools4U-deelnemers vergeleken met Controlegroep 2



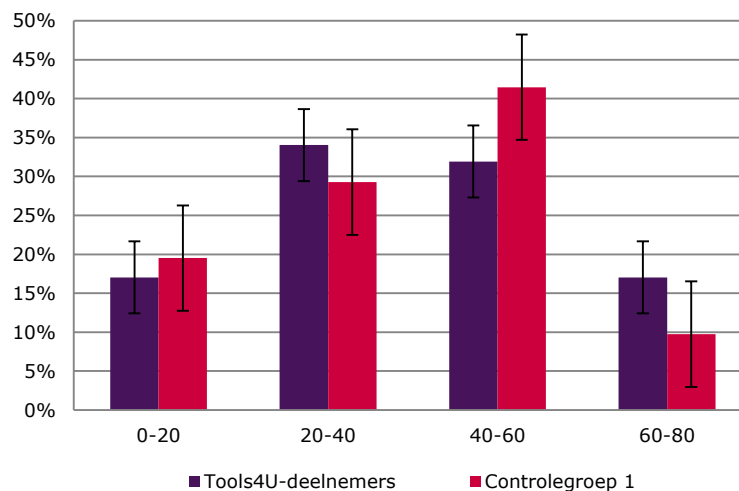
Noot: Log-rank = 0,10, $p=0,75$; Wilcoxon = 0,02, $p=0,89$; Tarone-Ware = 0,05, $p=0,82$.

4.2 Ernst eerste recidivezaak

Naast het percentage personen dat binnen drie jaar opnieuw een strafzaak heeft, is tevens gekeken naar de ernst van die strafzaak. De ernst van de delicten is gemeten met behulp van de zogeheten MSI, een maat die is gebaseerd op de hoogte en/of duur van de opgelegde straffen.

Binnen drie jaar hebben 47 jeugdigen van de Tools4U-groep en 41 jeugdigen uit Controlegroep 1 een recidivezaak. De gemiddelde MSI-score van het eerste recidivecontact van de Tools4U-deelnemers bedraagt 38,1. Het eerste recidivecontact van Controlegroep 1 heeft een vergelijkbare MSI-score, namelijk 37,5. De MSI-scores verschillen dan ook niet statistisch significant van elkaar ($t_{(df=86)}=-0,12$, $p=0,91$). Ondanks dat de gemiddelde MSI-score van de groepen niet verschilt, is niet meteen zeker dat geen van beide groepen relatief veel lichte of juist ernstige delicten heeft gepleegd. In figuur 4 vergelijken we daarom ook de percentages jeugdigen in de Tools4Ugroep en Controlegroep 1 die een recidivezaak met een MSI-score hebben van 0-20, 20-40, 40-60 of 60-80.²¹ De onderzoeksgroepen lijken op de laatste twee categorieën het meeste van elkaar te verschillen. Een groter gedeelte van de jeugdigen uit Controlegroep 1 heeft een MSI-score van 40 t/m 60. De Tools4U-deelnemers daarentegen hebben relatief vaker een recidivezaak met een score van 60 t/m 80. Deze verschillen zijn echter niet statistisch significant ($\chi^2_{(df=3)}=1,63$, $p=0,65$). Dus net zoals het aantal personen dat recidiveert, vinden we ook geen verschillen tussen de Tools4U-deelnemers en Controlegroep 1 in de ernst van de recidive.

Figuur 4 Ernst eerste recidivezaak (MSI-score) Tools4U-deelnemers vergeleken met Controlegroep 1

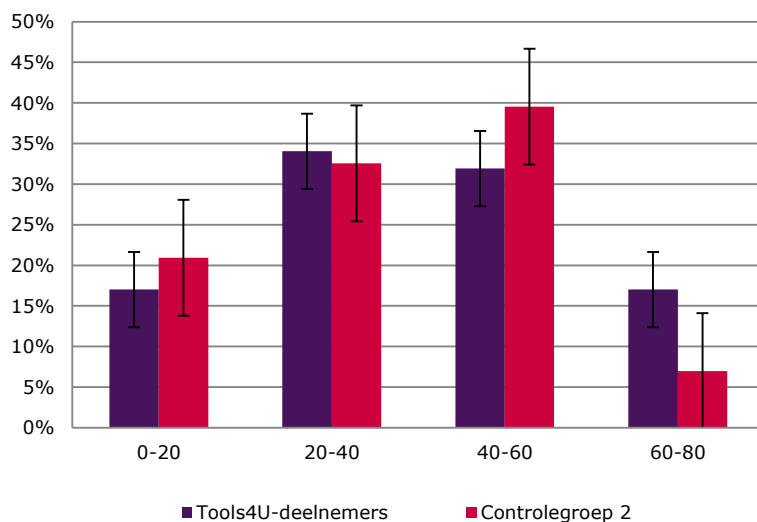


We hebben vervolgens ook de ernst van de recidivezaken van jongeren in Controlegroep 2 vergeleken met die van de Tools4U-deelnemers. Binnen drie jaar bleken 43 personen in Controlegroep 2 te hebben gerecidiveerd. De gemiddelde MSI-score van hun recidivezaak bedraagt 35,3. Dit wijkt niet statistisch significant af van de MSI-score van de Tools4U-deelnemers ($t_{(df=88)}=-0,62$, $p=0,54$). In figuur 5 staan de percentages jeugdigen beschreven die een MSI-score hebben van 0-20, 20-40, 40-60

²¹ Geen van de eerste recidivezaken in dit onderzoek had een MSI-score hoger dan 80.

of 60-80. Het grootste verschil tussen de Tools4U-deelnemers en Controlegroep 2 is te vinden in de laatste categorie. De Tools4U-deelnemers lijken oververtegenwoordigd wat betreft het aantal zeer hoge MSI-scores. De verschillen op de totale categorie-indeling zijn echter niet statistisch significant ($\chi^2_{(df=3)}=2,42, p=0,50$).

Figuur 5 Ernst eerste recidivezaak (MSI-score) Tools4U-deelnemers vergeleken met Controlegroep 2



4.3 Kenmerken die van invloed zijn op effectiviteit

Als laatste wordt nagegaan of de deelname aan Tools4U voor jeugdigen met bepaalde kenmerken een verschillend effect heeft gehad. Dit is alleen gedaan voor Controlegroep 1, aangezien we vooral geïnteresseerd zijn of er indicaties zijn dat Tools4U effectief is voor jeugdigen waarbij vooraf bepaalde vaardigheidstekorten gemeten zijn. In tabel 3 staan de uitsplitsingskenmerken opgesomd die wenselijk zijn in een moderatoranalyse mee te nemen om te zien of er een interactie is tussen het wel of niet deelnemen aan Tools4U en deze kenmerken. Op zich zouden we alle kenmerken in tabel 3 en interacties tussen de moderator en de achtergrondkenmerken en vaardigheidstekorten tegelijkertijd in één moderator analyse willen meenemen zodat rekening kan worden gehouden met onderlinge samenhang tussen de variabelen. Het aantal variabelen is echter te groot in verhouding tot het aantal personen die we in deze analyse kunnen meenemen ($n=222$). Om deze reden is met een *L1-penalized* regressieanalyse een selectie van de variabelen gemaakt. Kort samengevat houdt deze analyse in dat de coëfficiënten van variabelen die het minste bijdragen aan de voorspelling - wel of niet recidiveren- in de analyse door een 'penalty' gelijk worden gemaakt aan 0. Zie de methode en bijlage 2 voor een uitgebreide beschrijving van de procedure.

Tabel 3 Uitsplitsingskenmerken voor moderatoranalyse

Uitsplitsingskenmerken
Moderator variabele
Wel of niet deelnemen aan Tools4U
Achtergrondkenmerken
Sekse
Leeftijd
Herkomst
Type delict uitgangzaak
Aantal eerdere delicten
Vaardigheidstekorten
Motivatie voor behandeling
Cognitieve vertekeningen
Empathie
Coping stijlen
Sociale informatieverwerking

De L_1 -penalized regressieanalyse heeft acht kenmerken niet gelijk gemaakt aan nul en deze zijn meegenomen in de moderatoranalyse. Er is nog één kenmerk aan de moderatoranalyse toegevoegd zodat ook de 'hoofdvariabelen' waartussen de interactie is, mee worden genomen in het model. In tabel 4 staan de uitkomsten van de moderatoranalyse weergegeven.²²

Tabel 4 Moderatoranalyse (n=222)

Achtergrondkenmerken	exp (b)	95%-bi	z	p
Sekse (man)	2,06	1,39 - 3,06	3,6	0,00 *
Uitgangzaak: overige delicten ^a	1,82	0,96 - 3,43	1,8	0,07
Uitgangzaak: schoolverzuim	0,60	0,38 - 0,94	-2,2	0,03 *
Leeftijd eerste strafzaak	0,93	0,83 - 1,04	-1,3	0,18
Aantal eerdere delicten	1,42	0,94 - 2,15	1,7	0,10
Probleemscore 'anderen de schuld geven'	1,12	0,82 - 1,52	0,7	0,48
Probleemscore 'goed praten/verkeerd benoemen'	1,11	0,78 - 1,58	0,6	0,57
Controlegroep	0,45	0,17 - 1,25	-1,5	0,13
Controlegroep * probleemscore 'goed praten/verkeerd benoemen'	1,58	1,05 - 2,37	2,2	0,03 *

* $p < 0,05$

^a De overige delicten betreffen: verkeersdelicten (rijden onder invloed, joyriding), misdrijven op grond van de Opiumwet, misdrijven op grond van de Wet wapens en munitie en het beledigen van een ambtenaar in functie.

Uit de tabel blijkt dat drie variabelen statistisch significant samenhangen met de prevalentie van recidive. Dat zijn sekse, schoolverzuim als uitgangzaak en de interactie tussen het niet deelnemen aan Tools4U en denkfouten op het gebied van goedpraten of verkeerd benoemen.²³ Jongens bleken een ruim twee keer zo grote kans op recidive te hebben dan meisjes. Indien de jeugdige de straf opgelegd heeft

²² De fit van het model op basis van Harrell's C-statistiek is 0,71, dat wordt als voldoende gezien (Hosmer & Lemeshow, 2000). Deze maat berekent de kans dat de niet-recidivist een lagere hazard ratio heeft dan de recidivist plus de helft van de (mogelijk verwaarloosbare) kans dat de twee subjecten dezelfde hazard ratio's hebben. Ter vergelijking: een C-statistiek van 0,5 staat gelijk aan toeval; een C van 1 aan perfecte predictie.

²³ Goedpraten/verkeerd benoemen: antisociaal gedrag afschilderen als gedrag dat niet erg is, anderen geen schade toebrengt of zelfs acceptabel of wenselijk is; en/of in kleinerende of dehumaniserende termen over andere personen praten (Brugman et al., 2011).

gekregen in verband met schoolverzuim in plaats van het plegen van een delict dan was de recidivehazard kleiner. Tot slot blijkt er een interactie te zijn tussen de controlegroep en de mate van cognitieve vertekeningen op het gebied van goedpraten of verkeerd benoemen. De probleemscore op dit gebied blijkt vooral sterk positief samen te hangen met recidive indien de jeugdige deel uitmaakt van de controlegroep. De jeugdigen in de controlegroep lijken dus in grotere mate beïnvloed te worden door problemen op het gebied van goedpraten of verkeerd benoemen dan de Tools4U-deelnemers. Uit het onderzoek van Van der Stouwe en collega's (2015) is gebleken dat meer Tools4U-deelnemers zijn vooruitgegaan op dit probleemgebied dan jeugdigen uit de controlegroep. Dit verschil was marginaal significant ($p=0,059$). Het is dus mogelijk dat doordat bij een (net niet significant) groter gedeelte van de Tools4U-deelnemers na de straf de problemen op dit gebied zijn afgenomen, er een minder sterk verband is tussen de probleemscore voor aanvang van de straf en recidive. Geen van de andere vaardigheidstekorten liet echter een interactie-effect zien met het wel of niet deelnemen aan Tools4U.

5 Conclusie en discussie

5.1 Conclusie

In de huidige studie werd onderzocht in welke mate de leerstraf Tools4U effectief is in het voorkomen van recidive. Op basis van dit onderzoek kan niet geconcludeerd worden dat Tools4U meer of minder effectief is dan een andere leer- of werkstraf in het voorkomen van recidive. De recidive van Tools4U-deelnemers werd vergeleken met een tweetal controlegroepen: Controlegroep 1 die aan de Tools4U-groep gematcht werd op basis van demografische gegevens en sociale en cognitieve vaardigheden voorafgaand aan de straf en Controlegroep 2 die gematcht werd aan de Tools4U-groep op basis van demografische gegevens en gegevens over de delict-historie, zoals die beschikbaar waren in de OBJD. Voor de beide groepen werd er geen significant verschil in recidiveprevalentie gevonden met de Tools4U-groep. Ook wat betreft de ernst van het eerste recidivedelict werden geen significante verschillen gevonden.

Deze uitkomsten zijn onverwacht, omdat Tools4U voldoet aan bijna alle beginselen waaraan voldaan moet worden om effectief te kunnen zijn (*What Works*-beginselen, Andrews & Bonta, 2010a; Bonta & Andrews, 2007; Van der Laan, 2004). De uitkomsten zijn bovendien niet in overeenstemming met eerdere overzichtsstudies naar sociale-vaardigheidstrainingen en justitiële jeugdinterventies, die kleine langetermijneffecten vonden (Ang & Hughes, 2002; Cook et al., 2008; Lipsey et al., 2010; Lipsey, 2009). Er bestaan echter wel enige verschillen tussen de voorgaande overzichtsstudies en de huidige studie die het verschil in uitkomsten mogelijk verklaren.

Zo bestond in de overzichtsstudies naar sociale-vaardigheidstrainingen (Ang & Hughes, 2002; Cook et al., 2008) de langetermijnuitkomstmaat niet alleen uit recidive. Er werd gebruikgemaakt van alle uitkomstmaten die beschikbaar waren en daarom werden behalve recidive en antisociaal gedrag ook meer 'zachtere' maten van vaardigheden gemeten. Het is daardoor onduidelijk wat het effect van sociale-vaardigheidstrainingen alleen op recidive is. Het is goed mogelijk dat het kleine langetermijneffect in deze studies een effect op andere 'zachtere' uitkomstmaten weerspiegelt. Bovendien was de programma-inhoud en -integriteit voor de meeste van de geïncludeerde studies onduidelijk (Cook et al., 2008), waardoor het mogelijk is dat andere of aanvullende technieken en interventies ingezet zijn die tot effecten hebben geleid. Verder was de doelgroep in de geïncludeerde studies breder: de doelgroep was een bredere groep van jongeren met externaliserend probleemgedrag (niet per se delinquent). Het is daardoor onduidelijk in hoeverre Tools4U vergelijkbaar is met deze trainingen. Tot slot vergeleken de meeste onderzoeken het effect van de sociale-vaardigheidstraining met een controlegroep die geen behandeling kreeg. Het effect in deze studies is dan mogelijk mede het gevolg van het feit dat het beter is om iets te doen dan niets te doen (i.e., effect van aandacht) en zegt dan weinig over de superioriteit van sociale-vaardigheidstrainingen vergeleken met andere interventies. In de huidige studie daarentegen kregen de jeugdigen in de controlegroepen ook bij een werkstraf (minimale) begeleiding en was voorts onduidelijk in hoeverre zij ondertussen nog andere (voorwaardelijke) hulpverlening ontvangen hebben.

In de overzichtsstudies naar de effectiviteit van justitiële jeugdinterventies werden (bijna) alle bestaande interventies geïncludeerd (Lipsey et al., 2010; Lipsey, 2009). Binnen die veelvoud aan interventies waren sociale-vaardigheidstrainingen, samen met bijvoorbeeld cognitieve gedragstherapie, geïncludeerd onder de noemer vaar-

digheidstrainingen en beiden lieten kleine effecten op recidive zien. Ook hier geldt dat programma-inhoud en -integriteit vaak niet gemeten waren, dus dat het mogelijk is dat er andere technieken ingezet werden die tot effect hebben geleid.

Een andere verklaring voor de verschillen met eerdere overzichtsstudies (Lipsey et al., 2010; Lipsey, 2009) is dat justitiële jeugdinterventies het meest effectief waren bij de meest problematische jongeren en wanneer interventies multimodaal (aangrijpend op meerdere probleemgebieden tegelijk) ingezet worden. Omdat Tools4U zich richt op een relatief 'lichte' doelgroep en daarom weinig intensief is, met weinig ruimte voor multimodaal aangrijpen, mogen wellicht dus ook maar kleine effecten verwacht worden. Binnen overzichtsstudies is er voldoende power om deze kleine effecten te detecteren (zoals ook blijkt uit de eerdergenoemde studies), maar binnen een individuele studie, zoals de huidige, is er een veelvoud aan deelnemers nodig om de kleinste effecten te kunnen vinden.

Vergeleken met andere Nederlandse interventies zijn de huidige resultaten overigens minder onverwacht: er zijn in Nederland nog maar weinig interventies waarvan is aangetoond dat zij effectief zijn in het verminderen van recidive.

De huidige studie bood de mogelijkheid om te onderzoeken voor welke jongeren in het bijzonder de interventie het meeste effect heeft door middel van moderatoranalyses. Dit maakt het mogelijk om beter aan het responsiviteitsbeginsel (Andrews & Bonta, 2010a; Bonta & Andrews, 2007; Van der Laan, 2004) te voldoen en alleen jongeren die het meeste baat bij de interventie hebben te indiceren voor de interventie om daarmee de effectiviteit te vergroten. Uit deze moderatoranalyses bleek dat jongeren in de controlegroep die meer denkfouten met betrekking tot 'goedpraten/verkeerd benoemen' hadden vaker recidiveerden. Dit verband was minder sterk onder de Tools4U-deelnemers. Dit lijkt een aanwijzing te zijn dat Tools4U het recidiveverhogende effect van denkfouten op het gebied van goedpraten/verkeerd benoemen, geneutraliseerd heeft. Deze bevinding komt deels overeen met de uitkomsten van het onderzoek naar de effecten van Tools4U op de sociale en cognitieve vaardigheden: meer Tools4U-deelnemers waren vooruitgegaan op het probleemgebied 'goedpraten/verkeerd benoemen' dan jeugdigen uit de controlegroep, alleen dit verschil was slechts marginaal significant (Van der Stouwe et al., 2015). Uit de moderatoranalyse komen geen andere interactie-effecten naar voren.

Opvallend aan de moderatoranalyses is verder dat niet alle statische factoren waarvan bekend is dat die sterk voorspellend zijn voor recidive (Cottle, Lee & Heilbrun, 2001; Gendreau, Little & Goggin, 1996; Mulder, Brand, Bullens & Van Marle, 2011) ongeacht (behandel)conditie een effect laten zien. In de analyses blijkt, zoals geanticipeerd, dat jongens vaker recidiveren dan meisjes en dat spijbelaars minder vaak recidiveren (spijbelaars zijn de lichtste groep 'delictplegers'²⁴ in dit onderzoek). Gegevens over de delictsgeschiedenis van de jongere (leeftijd eerste strafzaak en het aantal eerdere veroordelingen) hadden echter geen significante invloed op de recidive. Een mogelijke oorzaak hiervan is dat de onderzoeksgroep weinig varieerde op deze kenmerken. Het merendeel van de deelnemers had geen of weinig eerdere veroordelingen en was 14 tot 16 jaar oud ten tijde van het eerste justitiecontact.

Tot slot lijkt de matching op basis van de OBDJ-data geslaagd. De vergelijking met de beide controlegroepen leidt tot dezelfde uitkomsten en er waren tussen de beide controlegroepen geen verschillen vooraf noch in recidive. De huidige studie heeft echter ook het belang van de eerste (meer tijdrovende) controlegroep aangetoond. Deze controlegroep maakte het mogelijk om behalve de invloed van statische risico-

²⁴ Spijbelen valt onder de zogenoemde statusdelicten, dat zijn wetsovertredingen die alleen gelden voor minderjarigen.

factoren (geslacht, leeftijd, herkomst en delict geschiedenis) ook de invloed van dynamische risicofactoren (sociale en cognitieve vaardigheden) te onderzoeken. Dit is in het bijzonder van belang omdat interventies zoals Tools4U zich richten op veranderingen in die factoren. Zo kan – ook in de toekomst – beter bepaald worden welke jongeren wel of juist niet toegeleid moeten worden naar welke interventie.

5.2 Discussie

De huidige studie kent enkele beperkingen. Een eerste beperking is dat het onderzoek niet voldoet aan de 'gouden standaard' van effectonderzoek, namelijk de willekeurige toewijzing van deelnemers aan de behandel- of controlegroep (Farrington, 2003). Het is daarom niet uit te sluiten dat het (gebrek aan) effect te wijten is aan de invloed van onbekende factoren die door de matching ontstaan zijn. Het huidige gebrek aan verschil in de beide matchingsmethoden lijkt echter de betrouwbaarheid van de matching te bevestigen. Een kanttekening hierbij is dat niet voor alle indicatiecriteria waarvoor jongeren Tools4U opgelegd kregen informatie beschikbaar was voor de onderzoeksgroepen. Het gaat bijvoorbeeld om de informatie uit het Landelijk Instrumentarium Jeugdstrafrechtketen (LIJ) dat wordt ingevuld door de RvdK wanneer een jongere in aanraking komt met justitie. Door deze informatie in het onderzoek te betrekken, kan wellicht de matching worden verbeterd en kunnen ook met deze criteria moderatoranalyses uitgevoerd worden. Een tweede beperking is het aantal deelnemers. Dit aantal was niet groot genoeg om kleine effecten op recidive of interactie-effecten te kunnen detecteren.

Voor de praktijk betekenen de huidige uitkomsten dat de effectiviteit van Tools4U minder groot of minder aannemelijk is dan gedacht werd. Er waren effecten op sociale en cognitieve vaardigheden meteen na de training, maar dit leidde niet aantoonbaar tot een effect op recidive op korte of langere termijn. De mate waarin Tools4U op basis van de eerdere onderzoeken voldoet aan de *What Works*-beginselelen bieden echter aanknopingspunten om de indicatie voor of uitvoering van Tools4U op bepaalde aspecten te verbeteren. De indicatiestelling kan bijvoorbeeld verbeterd worden door spijbelaars uit te sluiten van deelname, de betrokkenheid van ouders kan versterkt worden en er kan meer aandacht besteed worden aan het verbeteren van sociale en cognitieve vaardigheden, waar tot nu toe nog geen effecten op waren (zoals adequate copingstijlen, empathie en opvoedvaardigheden van de ouders, zie Van der Stouwe et al., 2013, 2015, 2016). Daarmee is de interventie beter onderbouwd, onderzocht en geëvalueerd dan de meeste andere (Nederlandse) interventies. Daarnaast is het mogelijk dat de korte termijneffecten op sociale en cognitieve vaardigheden (Van der Stouwe et al., 2015, 2016) op andere gebieden dan officiële recidive hun weerslag hebben. Het zou daarom zinvol zijn in een volgende studie tevens bredere uitkomstmaten te onderzoeken, zoals (zelfgerapporteerde) politiecontacten, dagbesteding, opleiding en werk, bij voorkeur bij een groter aantal deelnemers, zodat ook kleine(re) effecten gevonden kunnen worden.

Summary

Recidivism after Tools4U Comparative recidivism research Tools4U 2012

Tools4U is a social skills training imposed as a penal sanction aimed at increasing the social and cognitive skills of youth aged 12 to 18 years with a medium-sized risk of reoffending. Previous research into the effectiveness of this intervention has found positive results regarding social and cognitive skills. In addition, Tools4U is generally in line with the *What Works* principles, principles any intervention aiming to effectively reduce recidivism should comply to.

A previous study into the effects of Tools4U on recidivism compared 115 Tools4U participants to 108 youth with community service order a year later. Propensity score matching (PSM) was used to match these youth to the Tools4U participants. No significant differences were found between the Tools4U group and this control group after periods of 6 months, 12 months, the longest available period for all youth (1.46 years), or for the maximum observed period (2.76 years). Furthermore, no moderating effects were found for age, gender, ethnic background, or criminal history. A marginally significant effect was found for violent offences: if Tools4U participants reoffended they were less likely to commit a violent offence than youth in the control group. The current study extends this previous work, with its relatively short follow-up period, by examining recidivism over a longer time frame. Furthermore, the previous study made use of a control group recruited a year after the Tools4U group, resulting in a different observation period for the two groups. The current study makes use of an additional control group, made up of youth with a community service order during the same period as the Tools4U participants received their training order. In addition, analyses were carried out on the influence of social and cognitive skills measured at entry.

In the current study the following research questions are examined:

- To what extent is Tools4U effective in reducing recidivism? As measured by:
 - prevalence of recidivism;
 - seriousness of first new offence.
- Does the effectiveness of Tools4U depend on participant characteristics? Namely:
 - age, gender, ethnic background, criminal history;
 - skills deficiency measured at entry.

Method

This study uses data from several sources. The primary data source was compiled to research the effect of Tools4U on social and cognitive skills. Data from the Research and Policy Database for Judicial Information (in Dutch: OBJD) were linked to this data. Data from the OBJD was used to examine the criminal careers of the youth. The study covers 115 juvenile offenders who completed the Tools4U training in 2012. Recidivism by the Tools4U participants was compared to that of the two control groups. Control group 1 was compiled for a previous study into the effectiveness of Tools4U. Propensity score matching (PSM) was used to match the Tools4U participants to 108 juvenile offenders who were assigned community service or a different training order between 2013 and 2014. Control group 2 was also compiled using

PSM. For this group, the matching was carried out using the following ten background characteristics from the OBJD: gender, age at end of order, birth country, type of offence, sentencing district, modality (public prosecution office/juvenile court), seriousness of conviction offence, age at first conviction, number of previous convictions, and criminal case density. Control group 2 contains 115 juvenile offenders.

Recidivism was measured according to the regular methods of the WODC Recidivism Monitor. In this study, we examined the prevalence of recidivism, as well as the seriousness of the first new offence.

Results

Few differences were found between the background characteristics of the Tools4U participants and the control groups. Comparisons that did reach statistical significance were type of offence, age at first conviction, and sentencing district, between control group 1 and Tools4U participants.

Examining recidivism, we found that a year after completion of Tools4U 17% of participants had reoffended. After two years, this increased to 27% and after three years to 41%. Recidivism rates after one and two years for control group 1 were higher. After one year 25% of the youth in this group had reoffended and after two years 33% had reoffended. The rates for reoffending after three years were slightly lower for control group 1 than for the Tools4U participants, namely 38%. For control group 2 reoffending rates were 20% in the first year, 32% in the second year, and 38% in the third year. These differences in recidivism rates between the Tools4U participants and the two control groups were, however, not statistically significant.

The average seriousness of the first new offence by Tools4U participants was slightly higher than that for both control groups. However, these differences were not statistically significant. The results of the current study on recidivism, therefore, do not provide evidence for the effectiveness of Tools4U.

Finally, we examined whether the effect of participation in Tools4U differed according to participant characteristics. Only control group 1 was used for this analysis, as we did not have data on skills deficiencies for control group 2 and this was our specific area of interest. The results of the moderation analysis indicated that youth in the control group who demonstrated problems with minimizing or mislabelling at entry were more likely to reoffend than the Tools4U participants who demonstrated these deficiencies at entry. No significant effects were found for any of the other skill deficiencies or background characteristics when comparing the Tools4U participants with the control group.

Conclusion

Based on the results of this study we are unable to conclude that Tools4U is more or less effective than other training or community service orders in preventing recidivism. This result is surprising given that, based on previous research, Tools4U fulfils almost all the criteria necessary for an effective intervention. Furthermore, our findings are not in line with previous research into social skills training for juvenile offenders, which have found small positive long term effects. However, in these studies the program integrity was often not thoroughly measured and as a result other techniques used parallel to the intervention may have played a role. Furthermore, most previous studies have compared a group receiving social skills training

to a control group who received no treatment. The difference between the groups may, therefore, have simply been a consequence of receiving something rather than nothing, i.e., an 'attention' effect. Finally, Tools4U is not aimed at a very severe target group. It is consequently not very intensive, with little room for addressing multiple issues. Therefore, effect sizes are likely to be small and difficult to detect in a single study. In the Netherlands, few interventions have been shown to effectively reduce recidivism.

The one effect we did find was that youth in the control group who prior to the study demonstrated problems with minimizing and mislabelling were more likely to reoffend than Tools4U participants. This finding is in line with the results of a previous study into the effects of Tools4U on social and cognitive skills. That study found a marginally significant effect in this area, namely, that Tools4U participants made greater improvements in the problem areas of minimizing and mislabelling compared to the control group.

The results of this study need to be interpreted with caution. First, the participants were not randomly assigned to the Tools4U group. As a consequence, other unknown factors may have influenced the differences between the Tools4U participants and the control groups. Second, it is possible that the study did not have enough participants to be able to detect small but significant effects of the intervention. The results of this report mean that in practice the Tools4U intervention is likely less effective than has been previously considered. However, the extent to which Tools4U follows the *What Works* principles, based on previous research, does provide possibilities for improving the implementation of this intervention. In addition, future research is needed to examine broader outcome measures, such as (self-reported) police contact, daily routines, education, and employment. Such research would determine whether Tools4U is effective in making improvements in these areas. Studying a larger number of participants would also be preferable in order to detect small effects in these areas.

Literatuur

- Albrecht, G., & Spanjaard, H. (2011). *Tools4U: Training cognitieve en sociale vaardigheden als taakstraf: Handleiding voor trainers*. Duivendrecht: PI Research.
- Andrews, D.A., & Bonta, J. (2010a). *The psychology of criminal conduct* (5e ed.). New Providence, NJ: Matthew Bender & Company, Inc.
- Andrews, D.A., & Bonta, J. (2010b). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law*, 16, 39-55.
- Andrews, D.A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, 17(1), 19-52.
- Andrews, D.A., & Dowden, C. (2007). The risk-need-responsivity model of assessment and human service in prevention and corrections: Crime-prevention jurisprudence. *Canadian Journal of Criminology and Criminal Justice/La Revue Canadienne De Criminologie Et De Justice Pénale*, 49(4), 439-464.
- Ang, R.P., & Hughes, J.N. (2002). Differential benefits of skills training with anti-social youth based on group composition: A meta-analytic investigation. *School Psychology Review*, 31(2), 164-185.
- Asscher, J.J., Deković, M., Manders, W.A., Laan, P.H. van der, Prins, P.J.M., Arum, S. van, & Dutch MST Cost-Effectiveness Study Group. (2014). Sustainability of the effects of multisystemic therapy for juvenile delinquents in the Netherlands: Effects on delinquency and recidivism. *Journal of Experimental Criminology*, 10, 227-243.
- Beerhuizen, M.G.C.J., Wartna, B.S.J., Verweij, S., & Tollenaar, N. (2015). *De misdrijf-straf index: Op weg naar een maat voor de ernst van delicten afgeleid van de afdoening van strafzaken*. Den Haag: WODC. Memorandum 2015-3.
- Blossfeld, H.P. & Rohwer, G. (2002). *Techniques of event history modelling: New approaches to causal analysis* (2e ed.). Mahwah, NJ: Lawrence Erlbaum.
- Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation*. Ottawa, Ontario: Public Safety Canada. Cat. No. PS3-1/2007-6.
- Bosma, A.Q., Asscher, J.J., Laan, P.H. van der, & Stams, G.J.J.M. (2011). *Proces-evaluatie Tools4U*. Amsterdam: Kohnstamm Instituut.
- Breslow, N. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika*, 57(3), 579-594.
- Brugman, D., & Bink, M.D. (2011). Effects of the EQUIP peer intervention program on self-serving cognitive distortions and recidivism among delinquent male adolescents. *Psychology, Crime & Law*, 17, 345-358.
- Brugman, D., Nas, C.N., Velden, F. van der., Barriga, A.Q., Gibbs, J.C., Bud Potter, G., & Liau, A.K. (2011). *Hoe ik denk vragenlijst (HID): Handleiding*. Amsterdam: Boom test uitgevers.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cook, C.R., Gresham, F.M., Kern, L., Barreras, R.B., Thornton, S., & Crews, S.D. (2008). Social skills training for secondary students with emotional and/or behavioral disorders A review and analysis of the meta-analytic literature. *Journal of Emotional and Behavioral Disorders*, 16(3), 131-144.
- Cottle, C.C., Lee, R.J., & Heilbrun, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. *Criminal Justice and Behavior*, 28(3), 367-394.
- Cruise, K.R., Fernandez, K., McCoy, W.K., Guy, L.S., Colwell, L.H., & Douglas, T.R. (2008). The influence of psychosocial maturity on adolescent offenders' delinquent behavior. *Youth Violence and Juvenile Justice*, 6(2), 178-194.

- Dishion, T.J., Loeber, R., Stouthamer-Loeber, M., & Patterson, G. (1984). Skill deficits and male adolescent delinquency. *Journal of Abnormal Child Psychology*, 12(1), 37-53.
- Dodge, K.A., Price, J.M., Bachorowski, J., & Newman, J.P. (1990). Hostile attributional biases in severely aggressive adolescents. *Journal of Abnormal Psychology*, 99(4), 385-392.
- Farrington, D.P. (2003). A short history of randomized experiments in criminology. A meager feast. *Evaluation Review*, 27(3), 218-227.
- Freedman, B.J., Rosenthal, L., Donahoe Jr., C.P., Schlundt, D.G., & McFall, R.M. (1978). A social-behavioral analysis of skill deficits in delinquent and nondelinquent adolescent boys. *Journal of Consulting and Clinical Psychology*, 46(6), 1448-1462.
- Gaffney, L.R., & McFall, R.M. (1981). A comparison of social skills in delinquent and nondelinquent adolescent girls using a behavioral role-playing inventory. *Journal of Consulting and Clinical Psychology*, 49(6), 959-967.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34(4), 575-608.
- Goeman, J.J. (2010). L1 penalized estimation in the cox proportional hazards model. *Biometrical Journal*, 52(1), 70-84.
- Goense, P.B., Assink, M., Stams, G., Boendermaker, L., & Hoeve, M. (2016). Making 'what works' work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior. *Aggression and Violent Behavior*, 31, 106-115.
- Gresham, F.M., Cook, C.R., Crews, S.D., & Kern, L. (2004). Social skills training for children and youth with emotional and behavioral disorders: Validity considerations and future directions. *Behavioral Disorders*, 30(1), 32-46.
- Heer-de Lange, N.E. de, & Kalidien, S.N. (2014). *Criminaliteit en rechtshandhaving 2013: Ontwikkelingen en samenhangen*. Den Haag: Boom Lemma. Justitie in statistiek 4.
- Helmond, P.E., Overbeek, G., & Brugman, D. (2015). An examination of program integrity and recidivism of a cognitive-behavioral program for incarcerated youth in the Netherlands. *Psychology, Crime & Law*, 21(4), 330-346.
- Helmond, P.E., Overbeek, G., Brugman, D., & Gibbs, J.C. (2015). A meta-analysis on cognitive distortions and externalizing problem behavior: Associations, moderators, and treatment effectiveness. *Criminal Justice and Behavior*, 42(3), 245-262.
- Hoeve, M., Dubas, J.S., Eichelsheim, V.I., Laan, P.H. van der, Smeenk, W., & Gerris, J.R. (2009). The relationship between parenting and delinquency: A meta-analysis. *Journal of Abnormal Child Psychology*, 37(6), 749-775.
- Hosmer D.W., & Lemeshow S. (2000). *Applied Logistic Regression* (2e ed.). New York: John Wiley & Sons.
- Justitiële Informatiedienst (2009). *Justitieel Documentatie Systeem (JDS)*. Almelo: Justitiële Informatiedienst.
- Kaplan, E.L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282), 457-481.
- Kazdin, A.E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, 3(1), 1-27.
- Kazdin, A.E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, 63(3), 146-159.
- Kraemer, H.C., Frank, E., & Kupfer, D.J. (2006). Moderators of treatment outcomes: Clinical, research, and policy importance. *The Journal of the American Medical Association*, 296(10), 1286-1289.

- Kraemer, H.C., Wilson, G.T., Fairburn, C.G., & Agras, W.S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, 59(10), 877-883.
- Laak, J. ter, Goede, M. de, Aleva, L., Brugman, G., Leuven, M. van, & Hussmann, J. (2003). Incarcerated adolescent girls: Personality, social competence, and delinquency. *Adolescence*, 38(150), 251-265.
- Laan, P. van der (2004). Over straffen, effectiviteit en erkenning. *Justitiële verkenningen*, 30(5), 31-48.
- Larson, J.J., Whitton, S. W., Hauser, S.T., & Allen, J.P. (2007). Being close and being social: Peer ratings of distinct aspects of young adult social competence. *Journal of Personality Assessment*, 89(2), 136-148.
- Lipsey, M.W., Howell, J.C., Kelly, M.R., Chapman, G., & Carver, D. (2010). *Improving the effectiveness of juvenile justice programs: A new perspective on evidence-based practice*. Washington, DC: Center for Juvenile Justice Reform.
- Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders*, 4(2), 124-147.
- Lösel, F., & Farrington, D.P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventive Medicine*, 43(2, Supplement 1), S8-S23.
- Maag, J.W. (2006). Social skills training for students with emotional and behavioral disorders: A review of reviews. *Behavioral Disorders*, 32(1), 5-17.
- Mulder, E., Brand, E., Bullens, R., & Marle, H. van (2011). Risk factors for overall recidivism and severity of recidivism in serious juvenile offenders. *International Journal of Offender Therapy and Comparative Criminology*, 55(1), 118-135.
- Nas, C.N., Brugman, D., & Koops, W. (2005). Effects of the EQUIP programme on the moral judgement, cognitive distortions, and social skills of juvenile delinquents. *Psychology, Crime & Law*, 11(4), 421-434.
- Nas, C.N., Orobio De Castro, B., & Koops, W. (2005). Social information processing in delinquent adolescents. *Psychology, Crime & Law*, 11(4), 363-375.
- Paternoster, R., Pogarsky, G., & Zimmerman, G. (2011). Thoughtfully reflective decision making and the accumulation of capital: Bringing choice back in. *Journal of Quantitative Criminology*, 27(1), 1-26.
- R Core Team (2013). *R: A language and environment for statistical computing*. Wenen: R Foundation for Statistical Computing.
- Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Royston, P., & Parmar, M.K.B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21, 2175-2197.
- Spanjaard, H.J.M. (2013). *Tussentijdse beoordeling Tools4U: Notitie t.b.v. erkenningscommissie gedragsinterventies ministerie van veiligheid en justitie*. Duivendrecht: PI Research.
- Spanjaard, H.J.M., Regterschot-Von Lindheim, H., Groenhuijsen, L., & Bolt, R. (2012). *Stand van zaken Tools4U: Ten behoeve van beoordeling oktober 2012*. Utrecht/Duivendrecht: Raad voor de Kinderbescherming/PI Research.
- Smith, J.A., & Todd, P.E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators?. *Journal of Econometrics*, 125(1), 305-353.
- Stouwe, T. van der, Asscher, J.J., & Stams, G.J.J.M. (2013). *Doeltreffendheid Tools4U: Onderzoek naar de doelgroep, uitvoering en doeltreffendheid van leerstraf*. Amsterdam: Universiteit van Amsterdam, Faculteit der Maatschappij- en Gedragswetenschappen.

- Stouwe, T. van der, Asscher, J.J., Hoeve, M., Laan, P.H. van der, & Stams, G.J.J.M. (2016). Social skills training for juvenile delinquents: Post-treatment changes. *Journal of Experimental Criminology*, 12(4), 515-536.
- Stouwe, T. van der, Asscher, J.J., Hoeve, M., Laan, P.H. van der, & Stams, G.J.J.M. (2017). Follow-up of a social skills training (SST) for juvenile delinquents: Effects on reoffending. Submitted for Publication.
- Stouwe, T. van der, Asscher, J.J., Stams, G.J.J.M., Hoeve, M., & Laan, P.H. van der (2015). *Effectstudie Tools4U: Effecten op cognitieve en sociale vaardigheden. Onderzoek effecten van leerstraf Tools4U op cognitieve en sociale vaardigheden*. Amsterdam: Universiteit van Amsterdam, Forensische Orthopedagogiek.
- Tarone, R.E., & Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika*, 64(1), 156-160.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1), 267-288.
- Tweede Kamer der Staten-Generaal. (1996). Vergaderjaar 1995-1996: Taakstraffen. (Nota 2 nr. 24 807). 's-Gravenhage: Sdu Uitgevers.
- Veltri, C.O.C., Sellbom, M., Graham, J.R., Ben-Porath, Y., Forbey, J.D., & White, R.S. (2014). Distinguishing personality psychopathology five (PSY-5) characteristics associated with violent and nonviolent juvenile delinquency. *Journal of Personality Assessment*, 96(2), 158-165.
- Verweij, P.J.M., & Houwelingen, H.C. van (1994). Penalized likelihood in Cox Regression. *Statistics in Medicine*, 13, 2427-2436.
- Verweij, S., Tollenaar, N., Wartna, B.S.J., & Weijters, G. (2017). *Recidive tijdens en na de GBM: Vergelijkend recidiveonderzoek naar het effect van de gedragsbeïnvloedende maatregel uitgevoerd in de periode 2008-2013*. Den Haag: WODC. Cahier 2017-11.
- Wartna, B.S.J. (2009). *In de oude fout: Over het meten van recidive en het vaststellen van het succes van strafrechtelijke interventies*. Den Haag: Boom Juridische uitgevers.
- Wartna, B.S.J., Blom, M., & Tollenaar, N. (2011). *De WODC-Recidivemonitor: 4e herziene versie*. Den Haag: WODC. Memorandum 2011-3.
- Witten, D.M., & Tibshirani, R. (2010). Survival analysis with high-dimensional covariates. *Statistical Methods in Medical Research*, 19, 29-51.
- Zebel, S., Alberda, D.L., & Wartna, B.S.J. (2014). *Recidive na een reclasseringscontact: Overzicht en analyse van de terugval van personen na beëindiging van een toezicht of werkstraf in de periode 2002-2009*. Den Haag: Boom Lemma. Onderzoek en beleid 312.

Bijlage 1 Samenstelling Leescommissie

Voorzitter

Prof. dr. P.G.M. van der Heijden Universiteit Utrecht, University of Southampton

Leden

Drs. A.W.M. Eijken	Senior beleidsmedewerker, directie Sanctietoepassing en Jeugd, Ministerie van Justitie en Veiligheid
Mr. H.M.J. Ezendam	Coördinator Onderzoek, directie Beschermen, Aanpakken en Voorkomen, Ministerie van Justitie en Veiligheid
Drs. N. van Gemmert	Coördinerend Adviseur Analyse, Dienst Justitiële inrichtingen
Dr. A.W.M. van der Heijden	Parket Generaal
Drs. G. Mol	Informatie manager, Reclassering Nederland
Drs. M.M. Padmos	Senior beleidsmedewerker, directie Sanctietoepassing en Jeugd, Ministerie van Justitie en Veiligheid
Drs. A. Stolk	Procesmanager Erkende Gedragsinterventies, Raad voor de Kinderbescherming

Bijlage 2 Beschrijving analysetechnieken

Propensity score matching (PSM)

Voor de matching van Controlegroep 2 hebben we gebruikgemaakt van PSM (Rosenbaum & Rubin, 1983).²⁵ Een *propensity score* is een met behulp van regressie-analyse op de gemeten achtergrondkenmerken berekende waarschijnlijkheid dat een persoon aan een bepaalde conditie wordt toegewezen. In dit onderzoek is dat de kans dat men een Tools4U-training kreeg opgelegd. Een persoon in de Tools4U-deelnemersgroep wordt vergelijkbaar geacht met iemand in de controlegroep als zij dezelfde *propensity score* hebben, zelfs al verschillen ze onderling op bepaalde specifieke achtergrondkenmerken. Een gelijke *propensity score* houdt in dat voor een paar gemaakte personen, men aanneemt dat de toewijzing aan de conditie volledig willekeurig is, gegeven de gelijke *propensity score*.

Met behulp van *nearest neighbor*-matching is voor iedere jeugdige in de Tools4U-groep een persoon in de controlegroep geselecteerd. Dat betekent dat in de controlegroep de persoon geselecteerd wordt waarvan de *propensity score* het dichtst ligt bij die van de persoon in de behandelgroep (Smith & Todd, 2005).²⁶ We hebben gekozen voor een PSM-procedure zonder teruglegging. Een persoon kan dus niet meerdere keren aan een ander gemaakt worden. Als er twee of meer even goede matches zijn (zogenoemde *ties*), dan vond de selectie van de match willekeurig plaats. Om te checken of de PSM-procedure goed heeft gewerkt, werden de groepen voor en na matching op alle achtergrondkenmerken met elkaar vergeleken. De grootte van de verschillen wordt uitgedrukt in Cohens *d*. Cohen (1988) beschouwt een *d*-waarde van 0,2 als 'klein', 0,5 als 'middelgroot' en 0,8 als 'groot'. Daarnaast zijn *t*-toetsen (in het geval van continue variabelen) en Chi-kwadraattoetsen (in het geval van categorische variabelen) gebruikt om te toetsen of de verschillen tussen de onderzoeksgroepen statistisch significant zijn, tegen een alpha van 5%.

Survivalanalyse

Om het verloop van de prevalentie van recidive over de tijd te schatten is gebruikgemaakt van de productlimietschatter van Kaplan en Meier (1958). Dit is de standaardmethode om de recidivekans te schatten bij een dataset die gedeeltelijk *censored* is en waarbij de tijd continue gemeten is. Censoring wil zeggen dat de observatie van een persoon stopt. Dat kan zijn omdat diegene recidiveert of omdat de gegevensverzameling wordt beëindigd. Aangezien niet iedereen even lang gevolgd kan worden, treden er verschillen op in observatieduur. Door middel van survivalanalyse kan desondanks een betrouwbare schatting van de recidivekans worden verkregen. Voor het toetsen van verschillen in de recidiveprevalentie werden drie statistische toetsen gebruikt. Zo kon worden vastgesteld of verschillen tussen de survivalcurves van de verschillende onderzoeksgroepen zich over de gehele observatieperiode voordeden of alleen in een bepaald gedeelte daarvan. De eerste toets die we gebruikt hebben, is de log-ranktoets (Blossfeld & Rohwer, 2002). Deze toets is vooral gevoelig voor verschillen op het einde van de observatieperiode. De tweede toets is de gegeneraliseerde Wilcoxon-toets (Breslow, 1970). Deze toets is juist gevoelig voor verschillen aan het begin van de curve. De derde

²⁵ Voor meer informatie over de matching van Controlegroep 1, zie Van der Stouwe et al. (2015).

²⁶ In hoeverre het lukt om personen met vergelijkbare *propensity scores* te selecteren, wordt duidelijk in hoofdstuk 3.

toets, de Tarone-Waretoets, is ook net iets gevoeliger voor verschillen aan het begin van de curve, maar in mindere mate dan de Wilcoxontoets (Tarone & Ware, 1977).

Multipale regressieanalyse

Ten behoeve van de beantwoording van de derde onderzoeksvraag is een multipale regressieanalyse uitgevoerd. Een multipale regressieanalyse maakt het mogelijk een indicatie te krijgen van de unieke invloed van verschillende kenmerken op de recidive. We hebben hiervoor gebruikgemaakt van een Royston & Parmar regressiemodel (Royston & Parmar, 2002). Dit is een familie van parametrische survival-regressietechnieken die diverse *proportional hazards of proportional odds* modellen omvat. In deze survival modellen wordt in tegenstelling tot het Cox-regressiemodel meer aannames gedaan over de vorm van de *baseline hazard*,²⁷ maar minder dan bij volledig parametrische survivalregressie. Het voordeel van een vaste verdeling van de *baseline hazard* is dat dit stabielere resultaten kan opleveren.²⁸ Met behulp van het AIC-criterium is een keuze gemaakt voor het type verdeling en het aantal vrijheidsgraden voor de verdeling van de *baseline hazard*. De laagste AIC werd gevonden bij een *proportional odds* model met drie vrijheidsgraden. De effecten van het model worden uitgedrukt in een coëfficiënt, de exponent bèta (e^β , de *hazard ratio*). De *hazard ratio* is de verhouding tussen twee *hazards*: de kans om op een tijdstip te recidiveren en de cumulatieve kans om tot aan dat tijdstip niet te recidiveren. Bij categorische variabelen geeft de e^β aan in hoeverre de *hazard* op recidive toeneemt bij een bepaalde waarde van de variabele. Bij numerieke variabelen geeft de e^β aan hoeveel procent de *hazard* op recidive daalt of stijgt als de variabele met één eenheid toeneemt.

Penalized regressie

Omdat we in de multipale regressieanalyse ook de interactie-effecten tussen deelname aan Tools4U en achtergrondkenmerken/vaardigheidstekorten willen bekijken, zijn er in dit onderzoek veel verschillende variabelen beschikbaar, terwijl de onderzoeksgroep redelijk klein is. Daardoor ontstaat het gevaar dat het model te complex wordt en als het ware wordt overfit. In dat geval is de geschatte invloed van de achtergrondkenmerken op recidive niet meer betrouwbaar. Het is daarom van belang om het aantal variabelen in een multipale regressieanalyse niet te groot te maken. Met behulp van een *penalized* regressie-model²⁹ kan wordt gekeken welke set van variabelen de uitkomst het beste voorspellen. We hebben gebruikt gemaakt van de zogenoemde *L1-penalty*, λ_1 , ook wel *least absolute shrinkage and selection operator* (LASSO; Tibshirani, 1996) genoemd. Normaal gesproken worden in een regressiemodel de optimale coëfficiënten gevonden door de minus log-(partial)-likelijkheid te minimaliseren naar deze coëfficiënten. Bij de LASSO, wordt aan minus log-(partial)-likelijkheid tevens de som van de absolute waarden van de coëfficiënten maal een *penalty* toegevoegd. Door de toegevoegde laatste term maakt het model impliciet een afweging tussen het vergroten van de invloed van een variabele waardoor de *likelihood* toeneemt en een *penalty* op de grootte van de coëfficiënt van dezelfde variabele waardoor de *likelihood* afneemt. De coëfficiënten van variabelen

²⁷ De *baseline hazard* is de schatting van hoe het risico op recidive per tijdseenheid verandert, als alle covariaten in het model nul zouden zijn.

²⁸ De resultaten zijn stabielier omdat bij dit model, in tegenstelling tot Cox-regressie, alle observaties kunnen worden gebruikt voor het schatten van de *hazard ratio* en er maar een beperkt aantal parameters wordt gebruikt voor deze schatting.

²⁹ Hiervoor hebben we het packages *penalized* (Goeman, 2010) in de statistische programmeertaal R (R Core team, 2013) gebruikt.

die niet genoeg bijdragen aan de voorspelling van de uitkomst, krimpen tot ze gelijk zijn aan nul. Hoe hoger de *penalty* die wordt toegepast, hoe strenger de procedure en hoe minder variabelen uiteindelijk in het model worden opgenomen. De optimale waarde van de *penalty* wordt middels kruisvalidatie achterhaald, waarbij de *cross-validated likelihood* als criterium voor modelselectie wordt gebruikt (Verweij & Van Houwelingen, 1994). Een voordeel van deze methode (bijvoorbeeld ten opzichte van een *stepwise* regressiemodel) is dat de selectie van de variabelen niet vereist dat er meerdere modellen worden gefit; in een *penalized* regressie-model vindt de selectie van variabelen plaats tijdens het fitten van het model. Nadat met behulp van kruisvalidatie bepaald is hoe groot de *penalty*parameter moet zijn, wordt deze *penalty* op alle variabelen in het model toegepast. Er hoeft dus niet per individuele variabele een discretionaire beslissing te worden gemaakt of deze wel of niet in het model moet worden opgenomen (Witten & Tibshirani, 2010). Bovendien heeft de procedure geen last van numerieke problemen door multicollineariteit. Hierdoor is het mogelijk om alle achtergrondkenmerken en interactie-effecten tegelijkertijd in de procedure mee te nemen. De variabelen waarvan de coëfficiënten in de procedure niet naar nul 'gekrompen' zijn, zullen worden meegenomen in de uiteindelijke multipele regressieanalyses. Hun effect zal opnieuw geschat worden omdat de *penalty* ervoor zorgt dat ook de coëfficiënten die niet nul worden gekrompen, gebiast worden richting nul (ten behoeve van een verbeterde voorspelling). Aangezien we een zuivere schatting willen van het effect van de interventie en achtergrondkenmerken, moet deze zonder *penalty* geschat worden.